

Biosynthesis and mode of action of the β -lactone antibiotic obafluorin

Thomas Alexander Scott

John Innes Centre

Department of Molecular Microbiology

Norwich Research Park, Colney Lane, Norwich, NR4 7UH, UK

A thesis submitted to the University of East Anglia for the
degree of Doctor of Philosophy

September 2017

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Biosynthesis and mode of action of the β -lactone antibiotic obafluorin

Thomas Alexander Scott, John Innes Centre, September 2017

Abstract

DNA sequencing technologies have advanced rapidly in the 21st century and there are now an abundance of microbial genomes available to mine in search of novel biosynthetic gene clusters and the bioactive natural products they encode. Whilst an enormously exciting prospect, this abundance of genomic data presents a challenge in determining how to select clusters for further study.

To address this issue, this project focusses on the biosynthesis of the β -lactone antibiotic obafluorin produced by the soil bacterium *Pseudomonas fluorescens*. β -Lactones occur infrequently in nature but possess a variety of potent and valuable biological activities. They are commonly derived from β -hydroxy- α -amino acids, which are themselves privileged chiral building blocks in a variety of pharmacologically and agriculturally important natural products and medicines.

I report the delineation of the entire obafluorin biosynthetic pathway using complementary mutagenesis and biochemical assay-based approaches. As part of this work I have been able to characterise ObaG, a novel PLP-dependent L-threonine transaldolase responsible for the biosynthesis of an unusual nonproteinogenic β -hydroxy- α -amino acid precursor from which the β -lactone ring of obafluorin is derived. Phylogenetic analysis has shed light on the evolutionary origin of this rare enzyme family and has identified further gene clusters encoding putative L-threonine transaldolases. Furthermore, I have been able to biochemically assay an entire intact nonribosomal peptide synthetase that displays a noncanonical domain architecture and is responsible for obafluorin assembly and β -lactone ring formation.

These studies allowed both mechanism- and redundancy-guided genome mining strategies to be developed that might allow the specific targeting of novel chemistry in the uncharted reaches of the natural product world.

Acknowledgements

I would immediately like to express my enormous gratitude to Prof. Barrie Wilkinson for his invaluable guidance and support throughout this project. I have enjoyed countless discussions with him, both scientific and otherwise, and have benefitted immensely from his help and encouragement in conversations about my future. It has been both a tremendous honour and pleasure to be his first PhD student. I would also like to thank my supervisory committee members Dr. Jacob Malone and Prof. Mervyn Bibb FRS for their advice and careful scrutiny of my thesis work.

I also owe special thanks to Dr. Neil Holmes, who invested a great deal of his time in helping me learn molecular methods in my first year, and to Dr. Daniel Heine for his many contributions in the more chemistry-based elements of my research. I thoroughly enjoyed my time working with them both. Thanks also to Dr. Zhiwei Qin and Dr. Gerhard Saalbach for help in purifying and characterising obafluorin, and with proteomics analyses respectively. I am extremely grateful to Dr. Govind Chandra for helping me explore my ideas about new genome mining strategies and for his bioinformatics support on this project, and to Ms. Eposi Enjema Carine Solange for all her work investigating obafluorin bioactivity.

I feel extremely proud to have belonged to the Molecular Microbiology department at the John Innes Centre. It cultivates a unique research environment in which members readily share their experiences and support one another in their research in whatever way they can. My thesis would not have been possible without the help and support of numerous members of the department, both past and present: I think particularly of Dr. Silke Alt, Mr. Thomas Booth, Dr. Matt Bush, Dr. Juan-Pablo Gomez-Escribano, Dr. Richard Little and Dr. Javier Santos-Aberturas. I have really enjoyed my time in Mol. Micro. and have made many great friends. It will be with a very heavy heart that one day I move on to something new.

Finally, I would like to thank my family who have always been a source of unremitting support and my partner, Dhana Thomy, who has been constantly at my shoulder through both good times and more challenging times.

Author's declaration

The research described in this thesis was conducted entirely at the John Innes Centre between October 2013 and September 2017. All the data described are original and were obtained by the author, except where specific acknowledgement has been made. No part of this thesis has previously been submitted as for a degree at this or any other academic institution.

A significant proportion of this work has been the subject of the following publication, which is included in the appendices of this thesis:

Scott, T. A., Heine, D., Qin, Z. & Wilkinson, B.

An L-threonine transaldolase is required for L-*threo*- β -hydroxy- α -amino acid assembly during obafluorin biosynthesis.

Nat. Commun. **8**, 15935 (2017).

Abbreviations

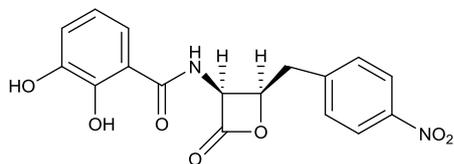
α-OH-β-AA	α -Hydroxy- β -amino acid
β-OH-α-AA	β -Hydroxy- α -amino acid
2,3-DHBA	2,3-Dihydroxybenzoic acid
2,4-DAPG	2,4-Diacetylphloroglucinol
3-HBA	3-Hydroxybenzoic acid
4-ABA	4-Aminobenzoic acid
4-APA	4-Aminophenylalanine
4-APP	4-Aminophenylpyruvate
4-HBA	4-Hydroxybenzoic acid
4-NPA	4-Nitrophenylacetaldehyde
4-NPAA	4-Nitrophenylacetic acid
4-NPE	4-Nitrophenylethanol
4-NPP	4-Nitrophenylpyruvate
AaRS	Aminoacyl-tRNA synthetase
ADC	4-Amino-4-deoxychorismate
ADH	Aldehyde dehydrogenase
AHAB	(2 <i>S</i> ,3 <i>R</i>)-2-Amino-3-hydroxy-4-(4'-aminophenyl)butanoate
AHBAS	3-Amino-5-hydroxybenzoic acid synthase
AHNB	(2 <i>S</i> ,3 <i>R</i>)-2-Amino-3-hydroxy-4-(4'-nitrophenyl)butanoate
AMP	Adenosine monophosphate
ANA	Anthranilic acid
ATP	Adenosine triphosphate
BA	Benzoic acid
BGC	Biosynthetic gene cluster
CLP	Cyclic lipopeptide
CoA	Coenzyme A
DAHPh	Deoxy-D-arabino-heptulosonate-7-phosphate
DNA	Deoxyribonucleic acid
DTT	DL-Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FAS	Fatty acid synthase
GCS	Glycine cleavage system
GHMT	Glycine hydroxymethyltransferase
GlyU	5'-C-glycyluridine
HCN	Hydrogen cyanide
HGT	Horizontal gene transfer
HPLC	High-performance liquid chromatography
HTS	High-throughput screening
IAA	Indole-3-acetic acid
IPM	Isopropylmalate
IPMS	Isopropylmalate synthase
IPP	Isopentenyl diphosphate
LCMS	Liquid chromatography-mass spectrometry
MGE	Mobile genetic element
MIC	Minimum inhibitory concentration
MS	Mass spectrometry

NAC	<i>N</i> -Acetylcysteamine
NADH	Nicotinamide adenine dinucleotide (reduced)
NMR	Nuclear magnetic resonance
NP	Natural product
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase
O/N	Overnight
PCS	Protein coding sequence
PK	Polyketide
PKS	Polyketide synthase
PLP	Pyridoxal-5'-phosphate
PPant	Phosphopantetheine
pPDC	Phenylpyruvate decarboxylase
PPi	Inorganic pyrophosphate
PQQ	Pyrroloquinolone quinone
PQS	<i>Pseudomonas</i> quinolone signal
PTM	Post-translational modification
QS	Quorum sensing
RiPP	Ribosomally synthesised and post-translationally modified peptide
RNA	Ribonucleic acid
SAL	Salicylic acid
SAM	S-adenosyl methionine
SHMT	Serine hydroxymethyltransferase
TA	Threonine aldolase
TEAB	Tetraethylammonium bicarbonate
ThDP	Thiamine diphosphate
THF	Tetrahydrofolate
ThrRS	Threonyl-tRNA synthetase
TTA	Threonine transaldolase

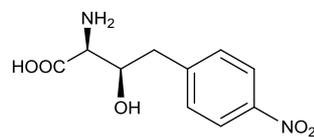
NRPS/PKS domains

A	Adenylation
ACP	Acyl carrier protein
ArCP	Aryl carrier protein
C	Condensation
Cy	Cyclisation
DH	Dehydratase
E	Epimerisation
ER	Enoyl reductase
KR	Ketoreductase
KS	Ketosynthase
PCP	Peptidyl carrier protein
TE	Thioesterase

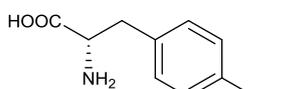
Compound Key



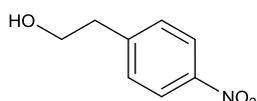
1
Obafluorin



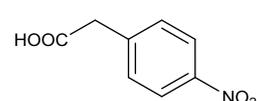
2
(2S,3R)-2-amino-3-hydroxy-4-(4'-nitrophenyl)butanoate
AHNB



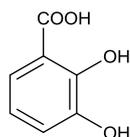
3
4-L-aminophenylalanine
4-APA



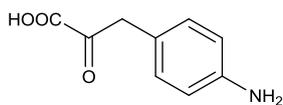
4
4-nitrophenylethanol
4-NPE



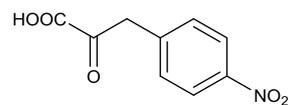
5
4-aminophenylacetic acid
4-NPAA



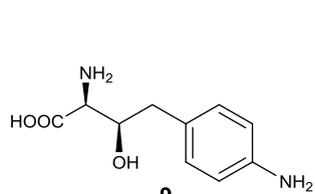
6
2,3-dihydroxybenzoic acid
2,3-DHBA



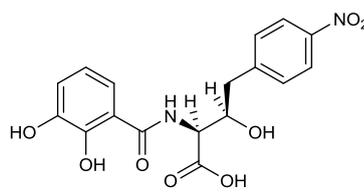
7
4-aminophenylpyruvate
4-APP



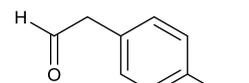
8
4-nitrophenylpyruvate
4-NPP



9
(2S,3R)-2-amino-3-hydroxy-4-(4'-aminophenyl)butanoate
4-AHAB



10
Hydrolysed 1



11
4-nitrophenylacetaldehyde
4-NPA

Table of contents

Abstract	i
Acknowledgments	ii
Author's declaration	iii
Abbreviations	iv
Compound Key	vi
Table of contents	vii
Index of Figures and Tables	xiii
Chapter 1: Introduction	1
1.1 Natural product discovery	2
1.1.1 What are natural products?.....	2
1.1.2 The rise of NP drug discovery.....	2
1.1.3 The fall of NP drug discovery.....	4
1.1.4 The second coming of NP drug discovery – the genome mining era.....	4
1.2 <i>Pseudomonas</i> spp. as producers of NPs	9
1.2.1 Overview of the pseudomonads and their genomes.....	9
1.2.2 The roles of NPs in <i>Pseudomonas</i> ecology.....	12
1.2.3 Biosynthesis of <i>Pseudomonas</i> NPs.....	14
1.2.3.1 Nonribosomal peptides (NRPs).....	15
1.2.3.2 Ribosomally synthesised and post-translationally modified peptides (RiPPs).....	17
1.2.3.3 Other amino acid-derived NPs.....	19
1.2.3.4 Polyketides (PKs) and fatty-acid-derived NPs.....	20
1.2.3.5 Hybrid PK/NRP NPs.....	24
1.2.3.6 Shikimic acid-derived NPs.....	25
1.2.4 Summary.....	25
1.3 Obafluorin	27
1.3.1 Discovery and characterisation.....	27
1.3.2 β -OH- α -AAs.....	28
1.3.2.1 Synthesis.....	28
1.3.2.2 β -OH- α -AAs in NP biosynthesis.....	30
1.3.2.3 Project outlook.....	31
1.3.3 β -Lactones.....	32
1.3.3.1 Ebelactones.....	32
1.3.3.2 Lipstatin.....	35

1.3.3.3 Salinosporamides and analogues.....	38
1.3.3.4 Oxazolomycin and analogues	39
1.3.3.5 Belactosins and cystargolides	43
1.3.3.6 Terpenoid β -lactones.....	45
1.3.3.7 Potential of synthetic β -lactones	45
1.3.3.8 Project outlook	46
1.3.4 Preliminary investigations of obafluorin (1) biosynthesis	46
1.3.4.1 Project outlook	50
1.4 Project objectives.....	52
Chapter 2: Materials and methods.....	53
2.1 General materials	54
2.2 Bacterial strains, plasmids and oligonucleotides.....	54
2.2.1 Bacterial strains	54
2.2.2 Plasmids	54
2.2.3 Oligonucleotides and primers.....	54
2.3 Culture media, buffers and solutions	57
2.3.1 Culture media	57
2.3.2 Buffers and solutions	58
2.3.3 Antibiotics	61
2.4 Cloning, screening and transformations	61
2.4.1 Preparation of <i>P. fluorescens</i> ATCC 39502 genomic DNA.....	61
2.4.2 Polymerase chain reaction (PCR).....	61
2.4.3 Agarose gel electrophoresis.....	62
2.4.4 Cloning of DNA fragments	63
2.4.5 Restriction site removal.....	63
2.4.6 Gibson assembly	64
2.4.7 Preparation and transformation of chemically competent <i>E. coli</i>	64
2.4.8 Plasmid purification and sequencing.....	65
2.5 Mutagenesis and complementation experiments	65
2.5.1 pTS1 and pJH10TS construct design.....	65
2.5.2 Conjugation of constructs into <i>P. fluorescens</i> ATCC 39502	65
2.5.3 pTS1 mutagenesis.....	66
2.5.4 Genetic and chemical complementation of Δ <i>oba</i> strains	66
2.5.5 Sample preparation for analysis of metabolite production	66
2.6 Expression and purification of recombinant proteins	67
2.6.1 Protein expression and purification	67
2.6.2 Protein concentration quantification	67

2.6.3 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)	68
2.6.4 Confirmation of protein identity by gel purification, trypsin digestion and MALDI-TOF/TOF analysis	68
2.7 In vitro Oba enzyme assays	68
2.7.1 His ₆ ObaH decarboxylase activity assay	68
2.7.2 His ₆ -ObaG transaldolase activity assays, time course and kinetic analysis	69
2.7.3 Coupled His ₆ -ObaH and His ₆ -ObaG activity assay.....	69
2.7.4 His ₆ -ObaG L-penicillamine competition assay	70
2.7.5 His ₆ -ObaG sodium borohydride reduction assay.....	70
2.7.6 His ₆ -ObaH hydroxylamine-trapping assay	70
2.7.7 <i>E. coli</i> threonyl tRNA synthetase (ThrRS) obafluorin (1)-binding assay 70	
2.8 Metal-binding assays	71
2.8.1 Chrome azurol S (CAS) assay	71
2.8.2 4-(2-pyridylazo)-resorcinol (PAR) assay.....	71
2.9 Antibacterial activity assays	71
2.9.1 Broth dilution assay	71
2.9.2 Disc and agar diffusion assays	72
2.10 Chemical methods	72
2.10.1 Chemical synthesis and isolation of substrates	72
2.10.1.1 Obafluorin (1)	72
2.10.1.2 Ethyl 5-(4'-nitrobenzyl)-4,5-dihydrooxazole-4-carboxylate	73
2.10.1.3 (2 <i>SR</i> ,3 <i>RS</i>)-2-amino-3-hydroxy-4-(4'-nitrophenyl)butanoic acid (AHNB; 2).....	73
2.10.1.4 (2 <i>SR</i> ,3 <i>RS</i>)-2-amino-3-hydroxy-4-(4'-aminophenyl)butanoic acid (AHAB; 9).....	74
2.10.2 Analytical methods and instrumentation.....	74
2.11 Phylogenetic analysis	76
2.12 Software	77
Chapter 3: Identification and delineation of the obafluorin (1) BGC	79
3.1 Introduction	80
3.2 Sequencing and annotation of the <i>P. fluorescens</i> ATCC 39502 genome	80
3.3 Preliminary analysis of the <i>P. fluorescens</i> ATCC 39502 genome	82
3.4 Identifying BGCs in the <i>P. fluorescens</i> ATCC 39502 genome	85
3.5 In silico analysis of the obafluorin (1) BGC	90

3.5.1 Biosynthesis of 2,3-DHBA (6)	92
3.5.2 Biosynthesis of 4-APP (7)	92
3.5.3 Biosynthesis of AHNB (2)	95
3.5.4 Obafluorin (1) dipeptide assembly.....	95
3.5.5 Nitro group formation	96
3.5.6 Precursor supply	98
3.5.7 Regulation of obafluorin (1) biosynthesis	98
3.5.8 Unassigned PCSs.....	101
3.5.9 Identifying <i>oba</i> BGC boundaries	102
3.6 Discussion	102
Chapter 4: Mutational analysis of the <i>oba</i> BGC.....	105
4.1 Introduction	106
4.2 Analysis of obafluorin (1) production	107
4.3 Construction of vectors for mutational analysis.....	109
4.3.1 pTS1 suicide vector construction	109
4.3.2 pJH10TS complementation vector construction	110
4.4 Mutational analysis of <i>oba</i> loci.....	112
4.4.1 Disrupting the biosynthesis of 2,3-DHBA (6)	112
4.4.2 Disrupting the biosynthesis of 4-APP (7).....	115
4.4.3 Disrupting the biosynthesis of AHNB (2)	115
4.4.4 Disrupting obafluorin (1) dipeptide assembly	115
4.4.5 Disrupting nitro group formation	118
4.4.6 Disrupting precursor supply	118
4.5 Discussion	123
Chapter 5: Biochemical characterisation of AHNB (2) biosynthesis	129
5.1 Introduction	130
5.2 Biochemical analysis of His₆-ObaH activity	132
5.3 Biochemical analysis of His₆-ObaG activity	132
5.4 His₆-ObaH/His₆-ObaG coupled reaction experiment.....	135
5.5 Characterising the PLP-dependency of ObaG	139
5.6 Exploring the potential of His₆-ObaG for synthetic applications	141
5.7 Discussion	143
Chapter 6: Phylogenetic analysis of SHMTs, L-TAs and L-TTAs	150
6.1 Introduction	151
6.2 Phylogenetic analysis of SHMTs, L-TAs and L-TTAs	154
6.3 Discussion	161

Chapter 7: Biochemical characterisation of obafluorin (1) assembly and β-lactone ring formation	165
7.1 Introduction	166
7.2 Investigating His₆-Obal A domain specificity	170
7.3 Characterising His₆-Obal TE-mediated β-lactone ring formation	173
7.4 Discussion	175
Chapter 8: Characterising the putative resistance gene and mode of action for obafluorin (1)	179
8.1 Introduction	180
8.2 Identifying the putative resistance gene in the <i>oba</i> BGC	183
8.3 Experimental validation of <i>obaO</i> as a resistance determinant of obafluorin (1)	185
8.3.1 Disruption of <i>obaO</i>	185
8.3.2 Assaying obafluorin (1) antibacterial activity	188
8.3.2.1 Broth microdilution assays	188
8.3.2.2 Disc diffusion assays	189
8.3.2.3 Agar diffusion assays	189
8.4 Identifying the obafluorin (1) mechanism of ThrRS inhibition	190
8.4.1 <i>In silico</i> comparison of putative obafluorin (1)- ‘sensitive’ and ‘resistant’ ThrRS amino acid sequences	190
8.4.2 <i>In silico</i> comparison of <i>obaO</i> with other putative ThrRS-inhibiting NP resistance determinants.....	191
8.4.3 <i>E. coli</i> ThrRS obafluorin (1)-binding assay	192
8.5 Investigating the metal-binding properties of obafluorin (1)	192
8.5.1 Chrome azurol S (CAS) assay to determine obafluorin (1) iron-binding potential.....	195
8.5.2 4-(2-pyridylazo)-resorcinol PAR assay to determine obafluorin (1) zinc-binding potential	195
8.6 Generation congeners of obafluorin (1) by mutasynthesis	195
8.7 Discussion	204
Chapter 9: Discussion	211
9.1 Summary of results	212
9.2 Future directions	213
9.2.1 Regulation of the biosynthesis of obafluorin (1).....	213
9.2.2 L-TTA directed evolution	214
9.2.3 Strategies for future genome mining efforts	216
9.2.4 Evolution of the <i>oba</i> BGC	220

9.2.5 Synthetic biological manipulations of the <i>oba</i> BGC	225
9.3 Concluding remarks	227
Bibliography	228
Appendix 1: Supplementary Figures and Tables	269
Appendix 2: Supplementary Sequence Data	289
Appendix 3: Publication	316

Index of Figures and Tables

Chapter 1

Figure 1.1	Example of an antibacterial bioassay plate	3
Figure 1.2	NPs from <i>Streptomyces coelicolor</i> A3(2)	6
Figure 1.3	The functional and environmental range of <i>Pseudomonas</i> spp.....	10
Figure 1.4	NPs from <i>Pseudomonas</i> spp.	13
Figure 1.5	Illustration of NRPS assembly line biosynthetic logic	16
Figure 1.6	Biosynthetic model for pyoverdine in <i>P. aeruginosa</i> PA01	18
Figure 1.7	Biosynthetic model and BGC for pseudomonine.....	18
Figure 1.8	Illustration of PKS assembly line biosynthetic logic.....	21
Figure 1.9	Biosynthetic model and BGC for mupirocin.....	23
Figure 1.10	Examples of β -OH- α -AA-comprising NPs	29
Figure 1.11	Reactions catalysed by SHMTs and TAs	29
Figure 1.12	Examples of β -lactone NPs.....	33
Figure 1.13	General mechanism of β -lactone acylation	33
Figure 1.14	Biosynthetic model and BGC for ebelactone.....	34
Figure 1.15	Illustration of β -lactone ring formation catalysed by OleC	36
Figure 1.16	Biosynthetic model for lipstatin	37
Figure 1.17	Biosynthetic model and BGC for salinosporamide A.....	40
Figure 1.18	Biosynthetic model and BGC for oxazolomycin.....	42
Figure 1.19	Biosynthetic models for belactosin C and cystargolide A	44
Figure 1.20	Illustration of interpreted results from D-[U- ¹³ C]glucose and L-[2,6- ³ H ₂]4-aminophenylalanine feeding experiments conducted by Herbert and Knaggs (1992a)	48
Figure 1.21	The tartronic semialdehyde pathway	49
Figure 1.22	Decarboxylative ThDP-dependent mechanism for the biosynthesis of AHNB (2) proposed by Herbert and Knaggs (1990)	49
Figure 1.23	Structures of amastatin and bestatin.....	51

Chapter 2

Table 2.1	Lab consumables used in this work	54
Table 2.2	Strains used in this work.....	57
Table 2.3	Plasmids used in this work.....	57
Table 2.4	Stock and working concentrations of antibiotics used in this work...	61
Table 2.5	High-fidelity PCR reaction mix and programme	62

Table 2.6	Colony PCR reaction mix and programme	62
Table 2.7	Software used for <i>in silico</i> analyses in this work	78
 Chapter 3		
Figure 3.1	Biosynthesis of the β -OH- α -AA AHNB (2) as proposed by Herbert and Knaggs (1990)	81
Figure 3.2	Illustration of the <i>P. fluorescens</i> ATCC 39502 genome	83
Figure 3.3	Assessment of genome synteny between <i>P. fluorescens</i> ATCC 39502 and three fully-sequenced model <i>P. fluorescens</i> strains	84
Figure 3.4	Functional assignment of PCSs in the genomes of four fully-sequenced <i>P. fluorescens</i> strains by the RAST server	86
Figure 3.5	BGC architecture and proposed biosynthetic pathway for the biosynthesis of obafluorin (1)	93
Figure 3.6	Proposed biosynthesis of 2,3-DHBA (6) and catechol-comprising NPs	94
Figure 3.7	Proposed biosynthesis of 4-APP (7) and NPs with 4-APP-derived moieties	94
Figure 3.8	<i>N</i> -oxygenation during NP biosynthesis	97
Figure 3.9	Role of DAHP synthase in the shikimic acid pathway to aromatic amino acids	99
Figure 3.10	Illustration of the reactions catalysed by the glycine cleavage system (GCS)	100
Table 3.1	Comparison of <i>P. fluorescens</i> ATCC 39502 genomic characteristics with those of other sequenced model <i>P. fluorescens</i> strains	85
Table 3.2	Summary of <i>P. fluorescens</i> ATCC 39502 antiSMASH output	87
Table 3.3	PCSs identified in, and flanking, the putative obafluorin (1) BGC....	92
 Chapter 4		
Figure 4.1	HPLC profile for a WT obafluorin (1) production culture extract at 270 nm.....	108
Figure 4.2	Vector maps for the suicide and complementation vectors constructed in this work	111
Figure 4.3	Examples of pick and patch plates to test for tetracycline sensitivity following pTS1-mediated gene disruption	111

Figure 4.4	BGC architecture and revised pathway for the biosynthesis of obafluorin (1) based on mutagenetic data	113
Figure 4.5	HPLC profiles for <i>obaJ</i> mutagenesis and complementation ` experiments at 270 nm	114
Figure 4.6	HPLC profiles for <i>obaL</i> mutagenesis and complementation experiments at 270 nm	114
Figure 4.7	Production culture flasks for WT, Δ and genetically complemented Δ strains.....	116
Figure 4.8	HPLC profiles for <i>obaF</i> mutagenesis and complementation experiments at 270 nm	116
Figure 4.9	HPLC profiles for <i>obaH</i> mutagenesis and complementation experiments at 270 nm	117
Figure 4.10	HPLC profiles for <i>obaI</i> mutagenesis and complementation experiments at 270 nm	117
Figure 4.11	HPLC profiles for <i>obaK</i> mutagenesis and complementation experiments at 270 nm	119
Figure 4.12	HPLC profiles for <i>obaC</i> mutagenesis and complementation experiments at 270 nm	119
Figure 4.13	HPLC profiles for <i>obaM</i> mutagenesis and complementation experiments at 270 nm	121
Figure 4.14	HPLC profiles for <i>obaG</i> mutagenesis and complementation experiments at 270 nm	122
Figure 4.15	First proposed biosynthetic mechanism for the biosynthesis of AHNB (2) involving ObaG acting as a transaldolase.....	126
 Chapter 5		
Figure 5.1	12 % SDS PAGE gel of His ₆ -ObaG and His ₆ -ObaH purification steps	133
Figure 5.2	Experimental characterisation of the biosynthesis of AHNB (2)	134
Figure 5.3	Characterisation of the biosynthesis of AHNB (2) by ObaG using ¹³ C NMR	136
Figure 5.4	LCMS profiles for <i>in vivo</i> [U- ¹³ C ₄ , ¹⁵ N]L-threonine feeding experiments	137
Figure 5.5	HPLC profiles for the His ₆ -ObaG reverse transaldolase assay with AHNB (2) and acetaldehyde	138
Figure 5.6	Proposed products of the His ₆ -ObaG transaldolase reaction using 4-NPP (8) and L-threonine as substrates	138

Figure 5.7	Phyre2 investigator analysis results for mouse SHMT (PDB: 1EJI)	140
Figure 5.8	Experimental confirmation that ObaG binds PLP	142
Figure 5.9	Time course and kinetic analysis of His ₆ -ObaG transaldolase activity	144
Figure 5.10	HPLC profiles for His ₆ -ObaG transaldolase assays with L-threonine and alternative aldehyde substrates	144
Figure 5.11	Proposed biochemical mechanism of ObaG L-TTA activity	146
Figure 5.12	L-TTA-catalysed reactions characterised in the literature and this work.....	148
Figure 5.13	NPs comprising GlyU moieties derived from LipK homologues.....	148
 Chapter 6		
Figure 6.1	Reactions catalysed by SHMTs, L-TAs and L-TTAs	153
Figure 6.2	Maximum likelihood tree of SHMTs, L-TAs and L-TTAs	156
Figure 6.2	Proposed biosynthetic pathways to alanylclavam, the clavamycins and valclavam.....	157
Figure 6.3	Proposed reactions catalysed by putative SHMTs/L-TAs in NP biosynthetic pathways.....	159
Figure 6.4	L-TTA-catalysed reactions characterised or proposed in the literature and this work	163
 Chapter 7		
Figure 7.1	Thioesterase-mediated PK/NRP chain release mechanisms	169
Figure 7.2	12 % SDS PAGE gel of His ₆ -Obal purification steps	169
Figure 7.3	NRPS A domain-catalysed reaction mechanism and substrate specificity assay principle.....	171
Figure 7.4	Hydroxylamine-trapping assay results	172
Figure 7.5	12 % SDS PAGE gel of His ₆ -ObaC and His ₆ -ObaK purification steps	174
Figure 7.6	Illustration of obafluorin (1)-SNAC formation	174
Figure 7.7	Illustration of the Obal A ₂ domain substrate-specifying residues...	176
 Chapter 8		
Figure 8.1	Illustration of the diverse bacterial targets for microbial NPs	181

Figure 8.2	Illustration of putative <i>oba</i> BGCs identified in genomic sequence data uploaded to GenBank.....	184
Figure 8.3	Mechanism by which tRNA synthetases (aaRSs) ligate amino acids to their cognate tRNAs.....	186
Figure 8.4	Mutagenesis HPLC data for Δ <i>obaO</i> and image of the corresponding production cultures	187
Figure 8.5	The interaction network of the Thr-AMS molecule with the Zn ²⁺ cofactor and the ThrRS active site residues.....	193
Figure 8.6	12 % SDS PAGE gel of <i>E. coli</i> His ₆ -ThrRS and His ₆ -ThrRS C480V purified protein sample serial dilutions	193
Figure 8.7	Mass spectra (74,000 – 79,000 <i>m/z</i> range) for <i>E. coli</i> ThrRS obafluorin (1)-binding assays.....	194
Figure 8.8	Preliminary qualitative liquid CAS assay results.....	196
Figure 8.9	Preliminary PAR assay spectrophotometric data	196
Figure 8.10	Illustration of the mutasynthetic strategy to generate novel congeners of obafluorin (1) in the Δ <i>obaL</i> strain.....	198
Figure 8.11	Image of mutasynthesis experiment production cultures.....	199
Figure 8.12	HPLC data for extracts of mutasynthesis experiment production cultures.....	200
Figure 8.13	HPLC and MS data for Δ <i>obaL</i> fed 0.4 mM benzoic acid (BA)	201
Figure 8.14	HPLC and MS data for Δ <i>obaL</i> fed 0.4 mM salicylic acid (SAL)	202
Figure 8.15	HPLC and MS data for Δ <i>obaL</i> fed 0.4 mM 3-hydroxybenzoic acid (3-HBA).....	203
Table 8.1	Obafluorin (1) disc diffusion assay results from Wells <i>et al.</i> (1982)	189
Table 8.2	Preliminary obafluorin (1) agar diffusion assay results	190
 Chapter 9		
Figure 9.1	Evolutionary origins of the <i>oba</i> BGC	223
 Appendix 1		
Supplementary Figure 1	Sequence alignment of authentic and putative SHMT amino acid sequences	270
Supplementary Figure 2	Sequence alignment of biochemically characterised L-TTAs.....	271

Supplementary Figure 3	Maximum likelihood tree of SHMTs, L-TAs and L-TTAs version 2	272
Supplementary Figure 4	Maximum likelihood tree of SHMTs, L-TAs and L-TTAs version 3	272
Supplementary Figure 5	Amino acid sequence alignment of NRPS TE domains	273
Supplementary Figure 6	Amino acid sequence alignment of authentic and putative ThrRSs with putative obafluorin (1) resistance proteins	276
Supplementary Figure 7	Sequence alignment of putative NP-resistant and <i>B.</i> <i>subtilis</i> ThrRS amino acid sequences	278
Supplementary Table 1	Oligonucleotides and primers used in this work	282
Supplementary Table 2	Source organisms and GenBank Accession numbers for SHMT, L-TA, L-TTA and AHBAS amino acid sequences used in phylogenetic analyses	288

Chapter 1:

Introduction

Chapter 1: Introduction

1.1 Natural product discovery

1.1.1 What are natural products?

Natural products (NPs), or specialised metabolites, represent an extensive family of chemical compounds that possess a wide variety of valuable bioactivities. They are molecules of bacterial, fungal, plant and marine animal origin that, whilst not essential for survival under laboratory conditions, offer the producing organism some selective advantage in the specific environmental conditions in which they are produced. The source of their wide-ranging bioactivities lies in their incredible chemical diversity and complexity; NPs can range from hundreds to thousands of Da in size, can contain unusual chemical elements (e.g. halogens) and many comprise one or more stereospecific carbon centres. NP research today encompasses a broad range of disciplines and expertise, driven by the need for novel pharmaceuticals to treat infections and diseases, and agrochemicals to improve the growth and protection of major crops to nourish the world's growing population.

1.1.2 The rise of NP drug discovery

Throughout the ages, humans have trusted in nature as a source of therapies for a whole range of infections and diseases. The oldest known record of natural medicines is the Egyptian "Ebers Papyrus" (ca. 1500 BCE) which documents over 700 drugs, mostly of plant origin (Borchardt, 2002). The major trigger in terms of modern NP discovery efforts was the serendipitous discovery of penicillin by Sir Alexander Fleming in 1928 (Fleming, 1929). It engendered Selman Waksman, a soil microbiologist at Rutgers University, to shift the efforts of his laboratory to search for natural compounds capable of killing bacteria known to cause human disease. By streaking bacteria and fungi isolated from the soil at right angles to pathogenic species, his group could observe the inhibition of pathogen growth by material diffusing from the producing organism (Figure 1.1), much as Fleming did. Waksman's systematic phenotypic screening efforts led to the discovery of Actinomycin (1940), Streptothricin (1941) and most notably Streptomycin (1943), the first compound active against *Mycobacterium tuberculosis* (Hopwood, 2007). Their successes prompted pharmaceutical companies and a small number of academic groups to begin their own strain isolation and screening programmes, and over the subsequent 30 years, a wealth of important NPs with antibacterial, antifungal, antiparasitic, anticancer and immunosuppressant properties were purified and characterised (Newman *et al.*



Figure 1.1. Example of an antibacterial bioassay plate. *Saccharopolyspora sp. KY21* (centre) is grown on nutrient agar and overlaid with agar containing *Bacillus subtilis*, representing a slight variation on the screens employed by Waksman in the 20th century. An antibacterial produced by the *Saccharopolyspora* strain is diffusing into the agar and inhibiting *B. subtilis* growth, creating a 'halo' of inhibition. This image is courtesy of Ms. Eleni Vikeli in the Wilkinson group.

2000; Demain 2014). This fruitful period is widely known as the 'Golden Age' of antibiotic discovery and many of the NPs discovered during this time, or their next-generation derivatives, are still in use today. These include, but are not limited to, the penicillins, cephalosporins, tetracyclines, aminoglycosides, polyenes, chloramphenicol, glycopeptides, macrolides and ansamycins. Infectious disease was the leading cause of death in humans in 1900, but over the course of the 20th century NPs revolutionised the practice of medicine, almost doubling the average human life span in the U.S. (Lederberg, 2000), a pertinent reflection of the impact of NPs discovered during the 1900s.

1.1.3 The fall of NP drug discovery

During the 1990s, combinatorial chemistry allowed for the rapid generation of thousands of unique synthetic compounds. With advanced robotics, target-based assays could be applied to screen entire libraries of these compounds numbering in the range of 5×10^5 to 4×10^6 – a process called high throughput screening (HTS) (Katz and Baltz, 2016). Unfortunately, NPs were not amenable to HTS strategies: obtaining a pure molecule from a complex culture mixture is often not trivial (especially at suitable concentrations for screening), compounds may act synergistically with activity lost following purification, and structural characterisation following purification can also be challenging (Li and Vederas, 2009). Compound rediscovery was a particular issue, given the rapid development of resistance to in-use antibiotics and the need to replace those becoming ineffective (Fischbach and Walsh, 2009; Davies and Davies, 2010). Advances in NMR and MS analyses could improve dereplication to a degree, but the volume and speed of HTS caused almost all pharmaceutical companies to abandon NP discovery programmes.

1.1.4 The second coming of NP drug discovery – the genome mining era

Despite initially being an irresistible prospect, HTS of large compound libraries failed to fulfil its potential in replacing NPs as a viable source of novel drug candidates. By 2002, the number of new chemical entities reaching the market had fallen to a two-decade minimum (Ortholand and Ganesan, 2004). A key limitation was the insufficient understanding of which areas of chemical space are best suited to interact with biological space. Compared to synthetic compounds, NPs cover a very distinct area of chemical space and normally comprise fewer nitrogen, halogen and sulphur atoms, contain more hydrogen bond donors and are considerably more oxygen rich. They also tend to possess more ring systems and a greater number of stereochemical features. NPs are the result of hundreds of millions of years of evolution, and their

complex chemical structures have been selected for precisely because of their ability to interact highly specifically with their biological targets. Furthermore, they also intrinsically possess the physiochemical properties necessary for gaining entry to cells, unlike many synthetic molecules (O'Shea and Moser, 2008). Their suitability as drugs is reflected in the fact that of the 1184 new chemical entities approved between 1981 and June 2006, more than half (52%) were NPs or NP-derived (Newman and Cragg, 2007), even with the withdrawal of NP discovery and screening programmes.

The understanding of the genetics and enzymology of NP biosynthesis thus continued to advance in the 1980s and 90s, stemming from pioneering work by Sir David Hopwood on the model NP-producing Actinomycete bacterium *Streptomyces coelicolor* (Hopwood, 2007). Rapid progress was also being made in the area of whole-genome sequencing, culminating in the first complete microbial genome for *Haemophilus influenzae* Rd being published in 1995 (Fleischmann *et al.*, 1995). This, combined with the crucial discovery that genes encoding entire biosynthetic pathways are found clustered together in the genomes of producing organisms (Malpartida and Hopwood, 1984), left NPs poised to make a return to the mainstage of drug discovery research efforts.

The sequencing of the *S. coelicolor* A3(2) genome in 2001 (Bentley *et al.*, 2002), in addition to microorganism genomes published subsequently (Ikeda *et al.*, 2003; Oliynyk *et al.*, 2007), led to a radical re-envisioning of the field of NP research. Not only could the known compounds produced by a particular microorganism be linked to their cognate biosynthetic gene cluster (BGC), but it was found that many more putative BGCs were encoded in their genomes beyond those for known compounds (Figure 1.2). Bacterial BGCs are rarely expressed constitutively as their products are energetically very costly to synthesise and are instead switched on in response to specific environmental cues or at different stages in microorganismal development (Bibb, 2005). Consequently, often only a fraction are expressed under standard laboratory culture conditions. As a result, as much as 90% of the chemical potential of these microorganisms had remained undiscovered by the traditional empirical screening strategies employed extensively during the 20th century.

This revelation established genome mining as an extremely powerful approach in NP discovery and led to the development of a range of new strategies for activating orphan or cryptic BGCs identified in genomic data (Challis, 2008; Rutledge and Challis, 2015; Zarins-Tutt *et al.*, 2015). General methods include varying culture

conditions, co-culture experiments, introducing chemical elicitors and epigenetic modifiers into cultures, as well as interfering with global regulators of transcription and translation. BGC sequence data allow for *in silico* predictions of final product structure to be made and often research groups are concerned with only a single promising or unusual biosynthetic locus. In these cases, pathway-specific approaches can be employed such as the over-expression of cluster-associated activators (or deletion of pathway-specific repressors), heterologous expression or *in vitro* reconstitution of pathways. Major advances have also been made in the field of metabolomics, allowing for faster dereplication and structural elucidation of isolated NPs (Li and Vederas, 2009; Ito and Masubuchi, 2014). More recently, molecular networking using tandem MS/MS data has become an increasingly popular tool for analysing complex mixtures of molecules in crude extracts. This allows improved novel NP identification and the detection of biosynthetic intermediates present in trace quantities (Watrous *et al.*, 2012; Hoffmann *et al.*, 2014; Crone *et al.*, 2016). Visualisation of the production of NPs *in situ* using Matrix Assisted Laser Desorption/Ionization (MALDI) – Imaging Mass Spectrometry (IMS) is also providing a greater understanding of the spatial and temporal distribution of NPs produced in different environmental and ecological contexts (Watrous and Dorrestein, 2011).

In the past decade, significant advances have been made in both the speed and accuracy of DNA sequencing, and the cost of sequencing per nucleotide has steadily declined (Gomez-Escribano *et al.*, 2016). To manage increasing amounts of genomic sequence data available, software and databases have been developed to automate BGC identification and annotation (Medema and Fischbach, 2015). One of the most widely used is antiSMASH (Blin *et al.*, 2017a; 2017b), which identifies BGCs based on the presence of biosynthetic enzymes putatively involved in NP biosynthesis; although they represent enormous chemical diversity, NPs are derived from a relatively modest set of building blocks and the biosynthetic machineries for their assembly are highly conserved. antiSMASH can identify >40 different BGC types in this way, in addition to making structural predictions of cluster products and showing homology to clusters in other sequenced genomes. Other software and databases (Medema and Fischbach, 2015) have been developed to catalogue characterised clusters and compounds (Pence and Williams, 2010; Medema *et al.*, 2015; Harborne, 2015;), to predict biosynthetic enzyme substrate specificity (Anand *et al.*, 2010; Röttig *et al.*, 2011), to connect genomic and mass spectrometric data (Medema *et al.*, 2014a; Wang *et al.*, 2016) and more recently, a retro-biosynthetic *in silico* analysis

platform has been created to further aid in NP discovery dereplication (Dejong *et al.*, 2016).

There are now >960,000 putative BGC entries in the Integrated Microbial Genomes Atlas of Biosynthetic gene Clusters (IMG-ABC), curated by the Joint Genome Institute (JGI) (Hadjithomas *et al.*, 2015; Ziemert *et al.*, 2016). Recent estimates based on the genomic data from 830 actinomycete genomes suggest that over 10^6 bacterial NPs remain to be discovered, with 10^4 - 10^5 of these possessing novel scaffolds (Doroghazi *et al.*, 2014). Given that >99% of bacteria are not readily culturable (Amann *et al.*, 1995), this number likely represents a considerable underestimate. The incredible wealth of chemical diversity potentially locked away in as yet uncultured organisms is reflected in the work of Piel and co-workers on *Entotheonella* factor, a sponge symbiont that produces a number of structurally distinct bioactive NPs (Wilson *et al.*, 2014). Advances in single-cell sequencing technologies and metagenomics approaches (Lasken, 2012; Wilson and Piel, 2013; Katz *et al.*, 2016), in addition to the development of innovative cultivation methods such as that which recently led to the discovery of teixobactin (Ling *et al.*, 2015), should make this uncharted chemical space increasingly accessible. Novel chemical structures are highly sought after as they may facilitate interaction with novel biological targets. In the antibiotic narrative, the inhibition of novel targets is critical in the battle against multi-drug resistant pathogenic bacteria. Considering the volume of genomic data now available, enriching for BGCs that give rise to novel chemical structures presents a key challenge in the field.

1.2 *Pseudomonas* spp. as producers of NPs

Historically, bacteria of the class *Actinobacteria* have been the principal source of NP derived-drugs (Bérdy, 2012), with more than half of all bacterial metabolites identified originating from members of the genus *Streptomyces* alone (Bérdy, 2005; Katz and Baltz, 2016). This is likely a reflection of their larger genome sizes encoding a greater average number of BGCs (5-10% of the genome) compared to other bacteria (Baltz, 2008), and the relative ease with which they can be cultured and studied in the laboratory. The rest of the eubacteria seem to be relatively poor producers, accounting for only 16% of all microbial NPs. Exceptions include *Burkholderia* (Liu and Cheng, 2014) and *Bacillus* spp. (Sansinenea and Ortiz, 2011), and more recently the cyanobacteria and myxobacteria are also proving to be rich sources of interesting NPs, accounting for 3.7 and 1.8% of known microbial NPs respectively (Bérdy, 2012). The ability of members of the genus *Pseudomonas* to produce a wide range of bioactive molecules is also well documented (Gross and Loper, 2009). Since obafluorin, the focus of this work, is a *P. fluorescens* NP, subsequent sections introduce the basic biology and genomics of this genus, and illustrate the significance of NPs in determining their ecology.

1.2.1 Overview of the pseudomonads and their genomes

Pseudomonas is a very diverse genus of gamma-proteobacteria comprising over 200 species (<http://www.bacterio.net>) that have evolved to occupy a diverse range of environments and niches (Figure 1.3). Many are saprophytic, playing important roles in the turnover of organic matter in the soil and are thus being investigated for their potential as bioremediators (Loh and Cao, 2008). Strains of *P. fluorescens*, *P. putida* and *P. protegens* colonise plants, where they support development through the biosynthesis of plant growth hormones and metabolites that suppress pathogens (Haas and Défago, 2005; Silby *et al.*, 2011; Loper *et al.*, 2012). Conversely, *P. syringae* strains are extremely successful plant pathogens, and can be assigned to one of over 50 pathovars based on host specificity (Buell *et al.*, 2003). Pathogenicity is not limited to plants, reflecting the opportunistic nature of these bacteria; *P. entomophila* orally infects and kill the larvae of many insect species (Vodovar *et al.*, 2006). *P. aeruginosa* is a clinical pathogen of patients with predisposing conditions such as cystic fibrosis or severe burns, and infections are notoriously difficult to treat due to a high degree of antibiotic resistance in this species (Lister *et al.*, 2009; Breidenstein *et al.*, 2011).

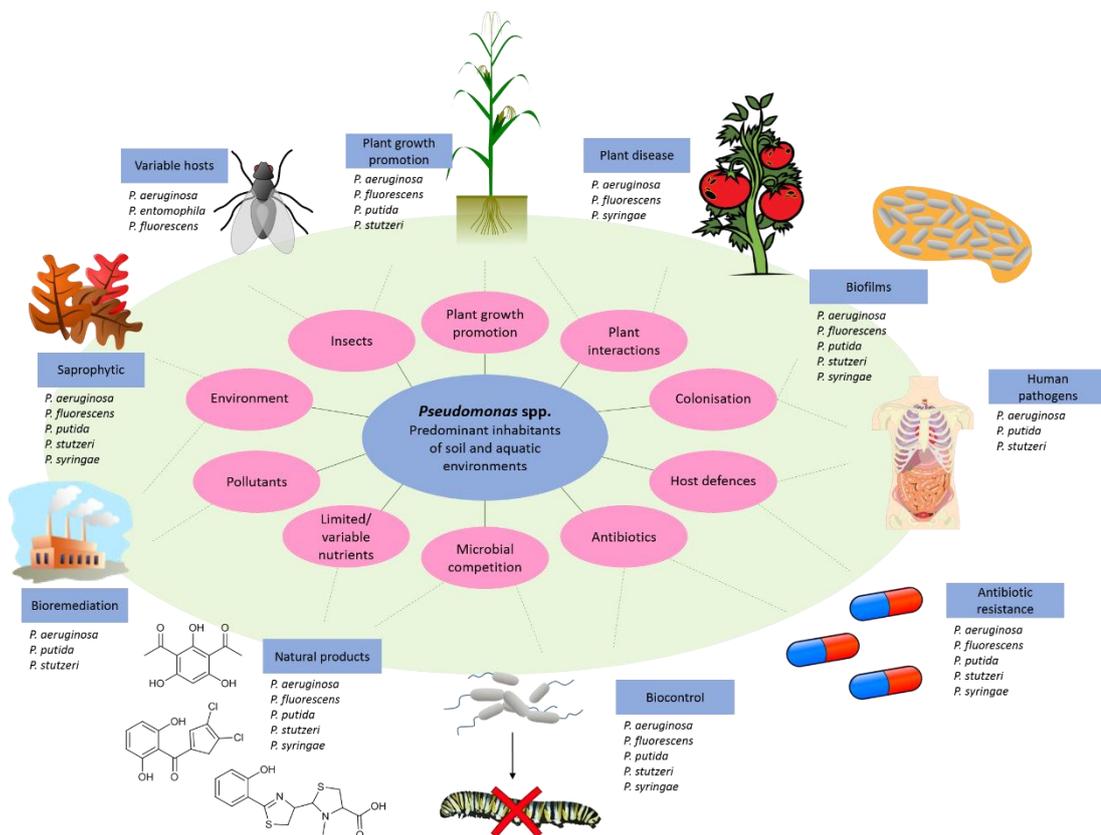


Figure 1.3. The functional and environmental range of *Pseudomonas* spp. The *Pseudomonas* common ancestor has encountered a wide range of abiotic and biotic environments that has led to the evolution of a multitude of traits and lifestyles with significant overlap among species. Silby et al. (2011) - Adapted with permission from Oxford University Press.

The incredible versatility and adaptability of *Pseudomonas* spp. is reflected in the enormous genetic diversity they exhibit. Only 25-35% of the core genome is shared by all members of the genus (Silby *et al.*, 2009; 2011; Baltrus *et al.*, 2011), and this typically comprises housekeeping genes and RNA genes essential for survival. Most genes are species-specific or shared by a particular sub-set of species, and make up what is known as the flexible genome. Compared to other bacteria pseudomonad genomes are large, typically 4.9-7.1 Mbp in size, which means there is a large proportion that is plastic and permissive to evolution. Genetic diversity is mediated in part by horizontal gene transfer (HGT) and the movement of mobile genetic elements (MGEs) such as phages, plasmids, transposons and genomic islands, that result in the acquisition or loss of genetic information. Whole-genome sequence alignments also show evidence of extensive intragenomic recombination events between repeat sequences. These are particularly abundant in *Pseudomonas* spp. and typically occur in the flexible genome, indicating a strong selection pressure for highly adaptive variable regions (Gross and Loper, 2009; Silby *et al.*, 2009; 2011). The core and flexible regions display a mosaic pattern of conserved and lineage-specific genes as the result of both processes (Spencer *et al.*, 2003; Linderberg *et al.*, 2008; Mathee *et al.*, 2008; Winstanley *et al.*, 2009).

The genomes of *Pseudomonas* spp. encode a large complement of genes that endow them with broad metabolic versatility, including complete sets for the pentose phosphate, Entner-Doudoroff and tricarboxylic acid central metabolic pathways (Stover *et al.*, 2000; Nelson *et al.*, 2002; Buell *et al.*, 2003; Paulsen *et al.*, 2005; Vodovar *et al.*, 2006). A sub-set are also able to metabolise plant-derived carbohydrates such as maltose, sucrose, trehalose and xylose, and even more complex molecules such as aromatics, long-chain fatty acids and hydrocarbons (Jiménez *et al.*, 2002; Martins dos Santos *et al.*, 2004; Paulsen *et al.*, 2005). Components for the catabolism of most amino acids are also present. Their genomes also encode numerous hydrolytic enzymes such as chitinases, proteases and lipases involved in the breakdown of soil polymers to liberate metabolic substrates (Loper *et al.*, 2012). Many species biosynthesise surfactants, solubilising agents that further increase the bioavailability of small molecules (Noordman and Janssen, 2002; Doong and Lei, 2003). Intuitively, *Pseudomonas* genomes encode many transport genes for the uptake of a diverse range of different nutrient sources (Buell *et al.*, 2003; Paulsen *et al.*, 2005). Extensive sets of regulatory genes are also encoded, reflecting the importance of directing gene expression in response to a variety of different environmental conditions and nutrient sources (Llamas *et al.*, 2014).

1.2.2 The roles of NPs in *Pseudomonas* ecology

High metabolic versatility and genome plasticity has created the ideal platform for the evolution of very diverse specialised metabolomes, which have played a critical role in the ability of *Pseudomonas* spp. to adapt to their diverse ecological niches. Their NP repertoires exhibit great chemical diversity (Figure 1.4), allowing the interaction with a broad range of biological targets. Many have antibiotic and antifungal activities, providing defence against competitors and predators. 2,4-diacetylphloroglucinol (2,4-DAPG) for example is produced by many plant-associated pseudomonads and is active against a wide range of plant-pathogenic fungi, most notably the take-all pathogen of wheat, *Gaeumannomyces graminis* var. *tritici* (Keel *et al.*, 1992; Raaijmakers and Weller, 1998). It also exhibits antibacterial and anthelmintic properties and triggers the systemic resistance of plants against disease (Nowak-Thompson *et al.*, 1994; Rezzonico *et al.*, 2007). Pyrrolnitrin, an inhibitor of fungal respiratory chains (Tripathi and Gottlieb, 1969), and hydrogen cyanide (HCN), an inhibitor of cytochrome C oxidase (Gross and Loper, 2009) are likewise commonly produced by plant growth-promoting species due to their disease suppressive activities. Some plant-associated species also produce phytohormones (Lindow *et al.*, 1998). Indole-3-acetic acid (IAA) influences many aspects of plant development and is produced at low concentrations by rhizosphere-inhabiting species to promote plant growth (Patten and Glick, 2002). Its effects are concentration-dependent however, and several plant-pathogenic species produce IAA at high concentrations as a virulence factor, causing hyperplasia (Smidt and Kosuge, 1978; Glickmann *et al.*, 1998; Spaepen *et al.*, 2007).

Pathogenic species produce many NP toxins and virulence factors to facilitate host infection. Phytotoxins such as tabtoxin and syringomycin from *P. syringae* cause chlorosis and necrosis symptoms in host plants for example (Bender *et al.*, 1999). Coronatine is a pseudomonad virulence factor that facilitates pathogen entry into the plant by causing stomata to open and which has also been shown to inhibit salicylic acid (SAL)-dependent plant host defences (Brooks *et al.*, 2005; Melotto *et al.*, 2006; Uppalapati *et al.*, 2007). Pederin is a highly cytotoxic agent produced by an as-yet uncultured symbiotic *Pseudomonas* spp. It was originally isolated from rove beetles of *Paederus* and *Paederidus* spp., the symbiont insect hosts, which use it as a chemical weapon for defence against predators (Kellner and Dettner, 1996; Piel, 2002).

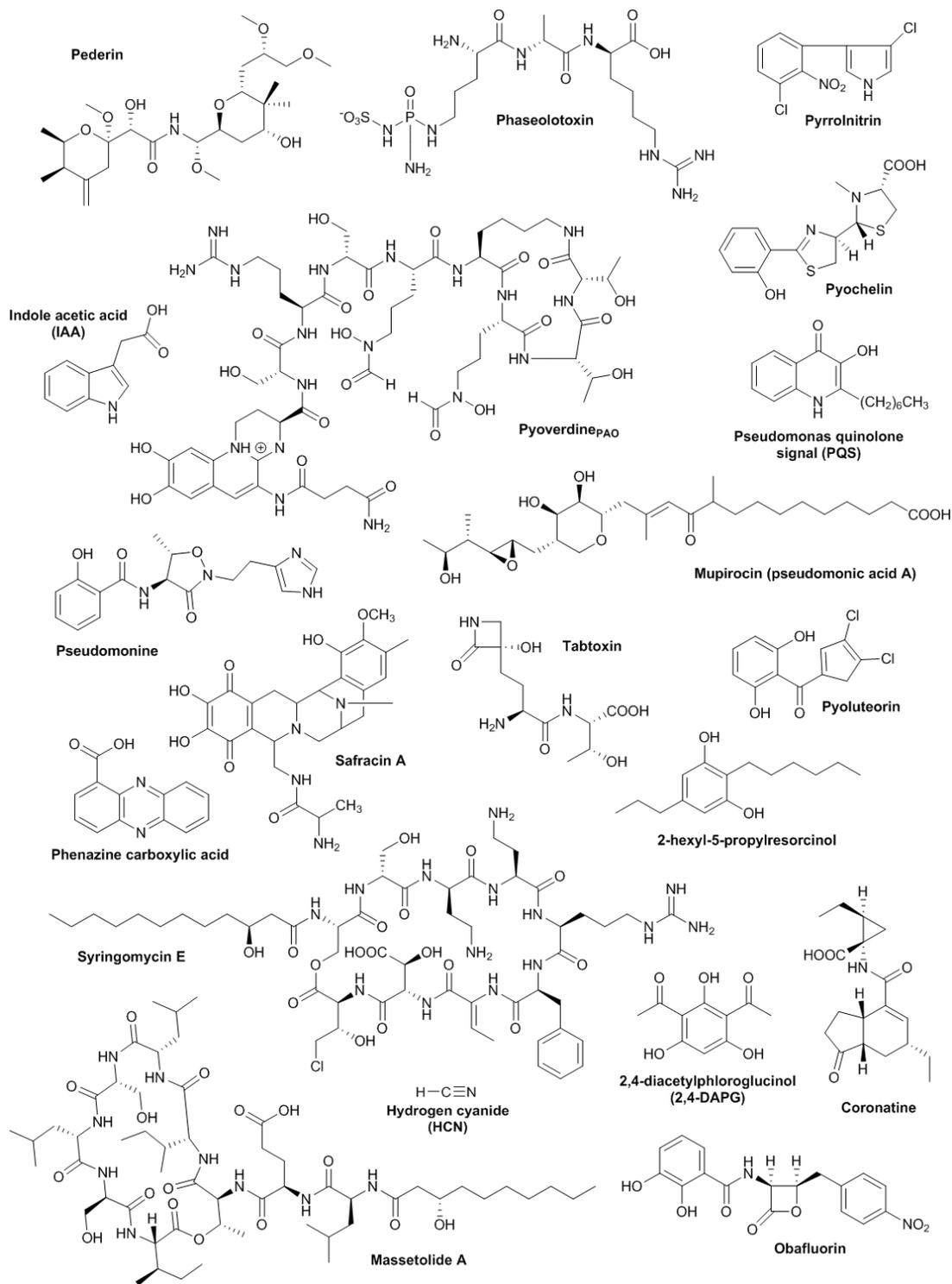


Figure 1.4. NPs from *Pseudomonas* spp.

NPs also play important physiological roles more akin to 'primary' metabolic functions. The phenazines produced by *P. aeruginosa* initially attracted attention for their broad-spectrum antibacterial, antitumour and antiparasitic bioactivities (Thomashow and Weller, 1988; Mavrodi *et al.*, 2006). However, their redox potentials are such that they can react extracellularly with higher-potential oxidants, effectively acting as electron shuttles between the cell and external substrates, allowing cells to continue to 'respire' under hypoxic conditions (Hernandez and Newman, 2001; Price-Whelan *et al.*, 2006). This is critical in biofilm formation where internal cells are likely to be starved of oxygen, and phenazine production is thus intrinsically linked to bacterial community morphogenesis (Dietrich *et al.*, 2013; Okegbe *et al.*, 2017). These redox-active metabolites also play critical roles as intercellular signals and influence transcriptional regulation by upregulating genes involved in stress response (Dietrich *et al.*, 2008).

Iron is essential for numerous metabolic and signalling functions, and is also an important enzyme cofactor for redox reactions (Wandersman and Delepelaire, 2004). However, it is insoluble in its oxidised form and highly toxic in its reduced form, and though abundant, is largely inaccessible in many environments. To overcome this issue, many *Pseudomonas* spp. secrete siderophores such as pyochelin (Cox *et al.*, 1981), pyoverdines (Visca *et al.*, 2006) and pseudomonine (Anthoni *et al.*, 1995). These chelate and solubilise ferric iron for uptake back into the producing-organism, whilst simultaneously depriving competitors of the same limited resource.

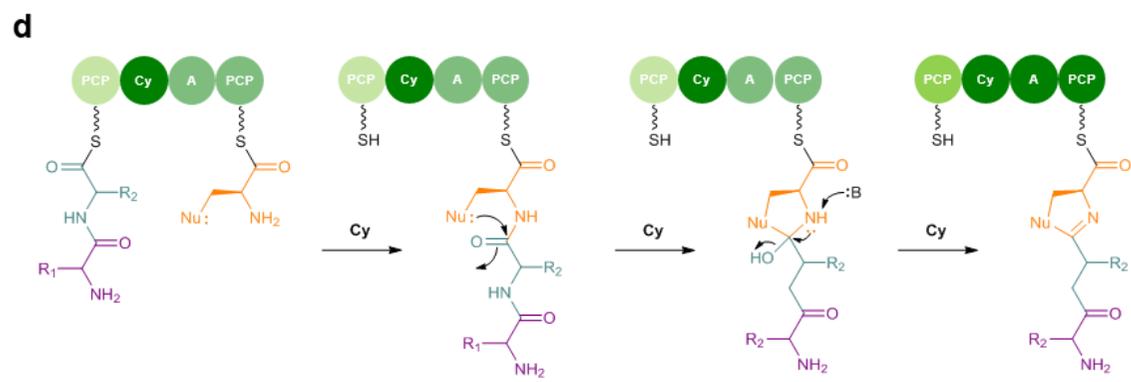
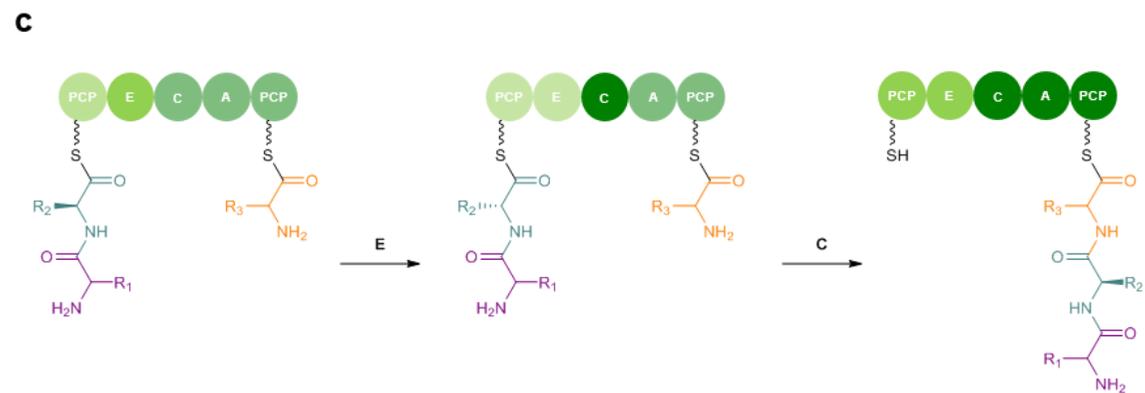
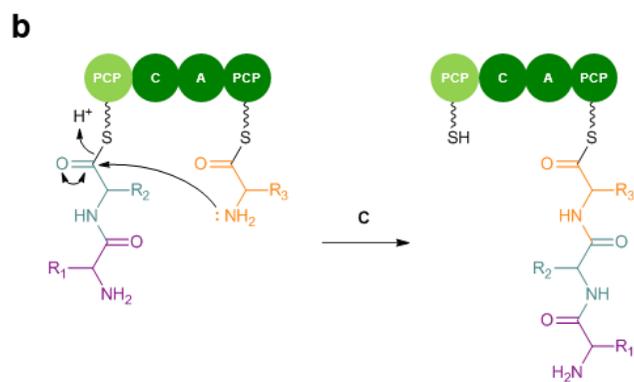
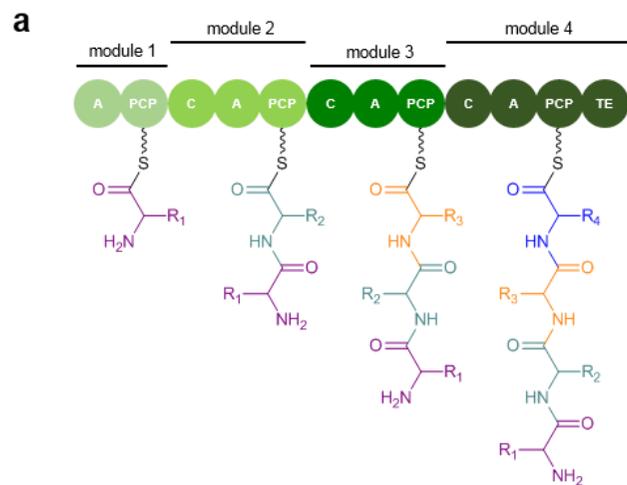
1.2.3 Biosynthesis of *Pseudomonas* NPs

The ability to produce a diverse set of chemical structures clearly provide the pseudomonads unique selective advantages that have allowed them to adapt to a broad range of ecological niches, but how is this chemical diversity achieved? To provide appropriate background knowledge for subsequent Introduction sections and Chapters, the major classes of NPs and their biosynthesis are here introduced using pseudomonad examples, which are illustrated in Figure 1.4. *Pseudomonas* spp. share much in common with the actinomycetes in that they utilise many of the same NP building blocks and machineries for assembling them. However, they also employ numerous unusual biosynthetic mechanisms to generate novel chemical structures and functionalities, and these will be highlighted in selected examples.

1.2.3.1 Nonribosomal peptides (NRPs)

Amino acids represent one of the principle building blocks used by *Pseudomonas* spp. in NP biosynthesis. They produce a broad array of structurally-diverse bioactive peptides which are assembled by large multifunctional proteins called nonribosomal peptide synthetases (NRPSs) (Finking and Marahiel, 2004; Fischbach and Walsh, 2006). These build peptides in a highly modular, assembly line-like fashion with each module responsible for catalysing the incorporation of a single specific amino acid into the growing peptide chain (Figure 1.5). The elongating peptide is shuttled from one module to the next, for the incorporation of each subsequent amino acid. Elongation modules comprise a minimal set of three domains: An adenylation (A) domain which recognises a specific amino acid substrate, activates it as an aminoacyl adenylate and tethers it to a phosphopantetheine (PPant) prosthetic group on the adjacent peptidyl carrier protein domain (PCP) via a thioester bond. Finally, a condensation (C) domain is then able to catalyse peptide bond formation between the PCP-tethered amino acid on the same module and the peptidyl intermediate bound to the PCP on the preceding module (Figure 1.5b). The assembly line is completed by an initial loading module, comprising A and PCP domains only, that initiates peptide assembly, and a termination module comprising a thioesterase (TE) domain, which catalyses release, and often cyclisation, of the final peptide product.

Chemical diversity in *Pseudomonas* peptide NPs is achieved by a multitude of different mechanisms. Beyond variation in peptide length and which amino acids are incorporated, independence from the ribosome has freed NRPSs to evolve to accept a broad range of nonproteinogenic amino acid and alternative acid substrates. The pyoverdines, for example, consist of a dihydroxyquinoline chromophore attached to a variable peptide chain of six to twelve amino acids, both of which comprise unusual acid substrates (Ravel and Cornelis, 2003 – Figure 1.6). Incorporation experiments revealed that 2,4-diaminobutyric acid is introduced by PvdL in chromophore assembly (Böckmann *et al.*, 1997), and PvdIJD subsequently extend the peptide by eight residues, incorporating two ornithine residues in the process. Further diversification is achieved by the presence of additional domains beyond the C-A-PCP paradigm that can further modify the peptide backbone during chain elongation. Epimerisation (E) domains for example allow the incorporation of D-configured amino acids (Figure 1.5c), and one D-tyrosine and a range of other D-configured residues are introduced into the *P. aeruginosa* PA01 pyoverdines by PvdL and PvdIJD E domains, respectively (Nowak-Thompson and Gould, 1994; Meyer, 2000; Ravel and Cornelis, 2003). Like many NRPs, pyoverdines are further modified post-assembly by the



actions of PvdA, an ornithine hydroxylase, and PvdF, a transformylase, which perform the respective hydroxylation and formylation at N-5 of ornithine, highlighting the many mechanisms by which chemical diversity can be generated among NP peptides.

Deviations from canonical NRP biosynthesis seem to be the rule rather than the exception for *Pseudomonas* NP peptides with many pathways involving additional NRPS domains and unusual organisations. Salicylic acid (SAL) derived from the shikimic acid pathway is incorporated as an atypical unit in the siderophores pseudomonine (Figure 1.7 - Mercado-Blanco *et al.*, 2001; Sattely and Walsh, 2008) and pyochelin (Serino *et al.*, 1997; Reimann *et al.*, 1998), two more siderophores that are both assembled by NRPSs. Both comprise NRPSs with cyclisation (Cy) domains (Figure 1.5d), which catalyse the formation of two oxazole rings from serine residues and one thiazole ring from a cysteine residue during pseudomonine and pyochelin biosynthesis respectively. PmsD in pseudomonine biosynthesis is an extremely unusual NRPS module that is composed solely of two Cy domains and one A domain (Mercado-Blanco *et al.*, 2001; Sattely and Walsh, 2008). Cy₂ transfers SAL to PmsG-tethered L-threonine, before Cy₁ catalyses cyclodehydration to generate a salicylmethyloxazolanyl-PmsG acylenzyme intermediate (Figure 1.7). The A domain activates another unusual precursor, *N*-hydroxyhistamine, which acts as a nucleophilic substrate for the PmsG C domain to release an oxazoline hydroxamate condensation product. The unusual organisation of this module serves to highlight the inherent plasticity of NRPS systems. Their highly modular repetitive organisations facilitate recombination events to yield novel assembly line organisations and novel products, and domains can sometimes evolve novel biosynthetic roles because of this process. Further biosynthetic flexibility is achieved in the substrate promiscuity of many NRPS A domains. PmsDEG were shown *in vitro* to accept 2,3-dihydroxybenzoic acid (2,3-DHBA) instead of SAL, or L-cysteine instead L-threonine, yielding the siderophores acinetobactin and anguibactin, respectively (Wuest *et al.*, 2009).

1.2.3.2 Ribosomally synthesised and post-translationally modified peptides (RiPPs)

Genome mining efforts of the 21st Century have revealed another major class of NP peptides - the ribosomally synthesised and post-translationally modified peptides (RiPPs) (Arnison *et al.*, 2013). They are encoded by small, structural genes which make them extremely difficult to identify by homology-based searches. However, the characterisation of different post-translational modification (PTM) enzymes involved

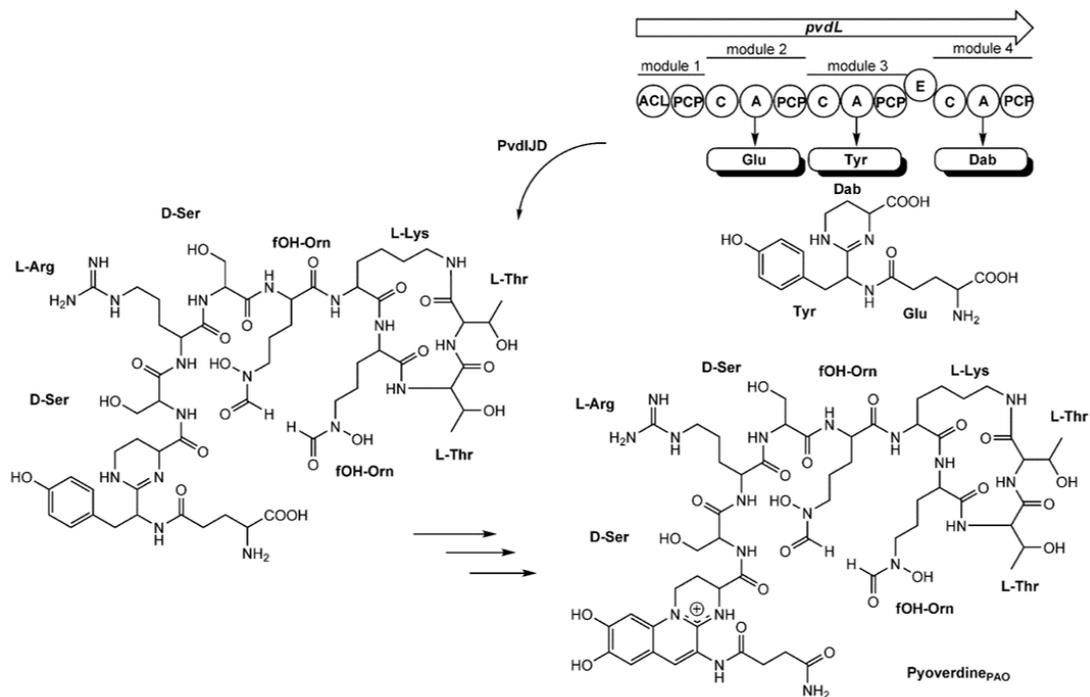


Figure 1.6. Biosynthetic model for pyoverdine in *P. aeruginosa* PA01. ACL = acyl-CoA ligase domain, Dab = 2,4-diaminobutyric acid. Gross and Loper (2009) - Reproduced with permission from The Royal Society of Chemistry.

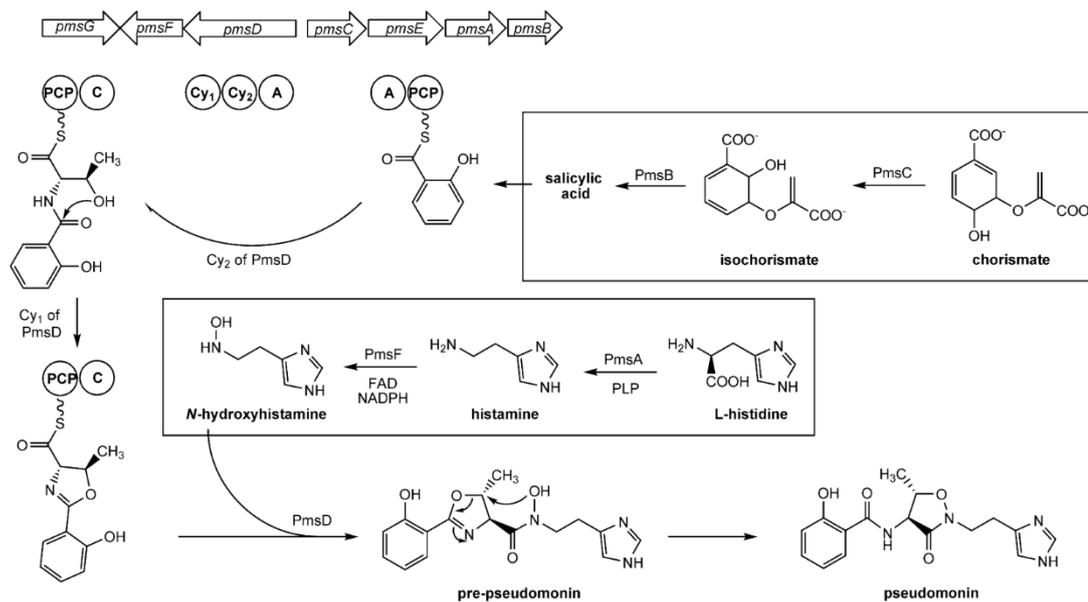


Figure 1.7. Biosynthesis model and BGC for pseudomonine. Gross and Loper (2009) - Reproduced with permission from The Royal Society of Chemistry.

in generating chemical diversity in this class of NP peptides is improving our ability to identify RIPP BGCs. Few examples have been characterised in *Pseudomonas* spp., although recently, microcin B-like BGCs have been identified in *P. syringae* and *P. putida* (Severinov *et al.*, 2007; Metelev *et al.*, 2013). Microcins were originally identified in *E. coli* and are peptides less than 10 kDa in size that contain multiple oxazole and thiazole rings and are potent inhibitors of DNA gyrase (Yorgey *et al.*, 1994; Li *et al.*, 1996; Duquesne *et al.*, 2007). Whilst *mcb*-like operons were identified in several *P. syringae* pathovars, their products have so far only been detected following heterologous expression in *E. coli* (Metelev *et al.*, 2013).

Genes encoding proteins of unknown function in the BGC for methanobactin biosynthesis, have allowed the identification of a homologous BGC in *P. fluorescens*, but it is yet to be characterised (Kenney and Rosenzweig, 2013). Methanobactins are a family of copper-binding RIPPs initially identified in the methanotroph *Methylosinus thrichosporium* OB3b, which requires copper as a cofactor for methane metabolism (DiSpirito *et al.*, 1998; Téllez *et al.*, 1998; Kim *et al.*, 2004). Advances in our understanding of RIPP biosynthesis will continue to inform genome mining strategies and will likely lead to the identification of novel pseudomonad RIPPs in the future.

1.2.3.3 Other amino acid-derived NPs

Phaseolotoxin, a phytotoxic tripeptide consisting of ornithine, alanine and homo-arginine connected to a sulfodiaminophosphinyl moiety, has been assigned to a 28 kb cluster devoid of NRPS genes (Peet *et al.*, 1986; Zhang *et al.*, 1993; Aguilera *et al.*, 2007). These building blocks are thought to be assembled by the L-amino acid ligase PhtU, encoded in the *P. syringae* pv. *phaseolicola* 1448A phaseolotoxin cluster (Arai and Kino, 2008), rather than by a canonical NRPS-based route. Other methods of ribosome independent peptide bond formation in NP biosynthesis are known (Goswami and Van Lanen, 2015), but have for a long time gone undetected by software such as antiSMASH.

The pseudomonads have evolved multiple routes to assemble amino acid building blocks in NPs, and several are even the result of a series of modifications of individual amino acids. The antifungal pyrrolnitrin is a chlorinated phenylpyrrole that is derived from multiple modifications of L-tryptophan, for example (Hammer *et al.*, 1997; Kirner *et al.*, 1998). It is the product of a four-gene locus *prnABCD* that catalyses two chlorinations, a decarboxylation (to form a phenylpyrrole from the indole) and an amino group *N*-oxygenation to install a nitro- functionality. IAA is also derived from

L-tryptophan and its biosynthesis can proceed *via* one of two different routes in *Pseudomonas* spp.: The indole-3-acetamide pathway (Costacurta and Vanderleyden, 1995) and the indole-3-pyruvic acid pathway (Patten and Glick, 1996). Further important examples of modified amino acid NPs include HCN which is derived from glycine (Wissing, 1974; Laville *et al.*, 1998; Blumer and Haas, 2000) and tabtoxin which has its origins in the L-lysine biosynthetic pathway (Unkefer *et al.*, 1987; Roth *et al.*, 1990; Liu and Shaw, 1997; Kinscherf and Willis, 2005).

1.2.3.4 Polyketides (PKs) and fatty acid-derived NPs

Acetyl-CoA and malonyl-CoA are also very common building blocks, used by pseudomonads and other microorganisms in the assembly of polyketide (PK) NPs. PKs are assembled by polyketide synthases (PKSs), which catalyse decarboxylative Claisen-like thioester condensation reactions to connect units together in an analogous manner to fatty acid synthases (FASs) (Staunton and Weissman, 2001; Fischbach and Walsh, 2006; Hertweck, 2009). Like NRPSs, PKSs are multifunctional proteins organised into modules, each responsible for the addition and modification of an extender unit to the growing PK chain (Figure 1.8a and b). PKS modules are also minimally composed of three domains: an acyltransferase (AT) domain selects an extender unit and tethers it to the PPant group of its cognate acyl carrier protein (ACP) domain via a covalent thioester bond. A ketosynthase (KS) domain then catalyses condensation between the extender unit and PK intermediate tethered to the preceding module ACP domain. Analogously to NRPSs, the loading module lacks a KS and the terminal domain on the terminal module is typically a TE domain for PKS product release.

Key distinctions from FASs are that PKSs can accept a greater variety of substrates, display a greater range of final product chain lengths, and exhibit varying levels of reduction following each round of chain elongation, where FASs fully reduce every acyl unit. This is due to the presence of three additional catalytic domains that are invariably present in FASs but 'optional' in PKSs: Ketoreductase (KR) domains convert the β -ketoacyl-S-ACP arising from KS-mediated condensation to a β -hydroxyacyl-S-ACP, dehydratase (DH) domains subsequently catalyse dehydration to form the α,β -enoyl-S-ACP, and finally enoylreductase (ER) domains reduce the conjugated olefin to the saturated acyl-S-ACP (Figure 1.8c). These additional domains are also able to alter NP stereochemistry. PKSs are categorised into three major types: type I are large modular proteins, type II consist of complexes of monofunctional proteins that use acyl-CoA substrates, and type III which are most

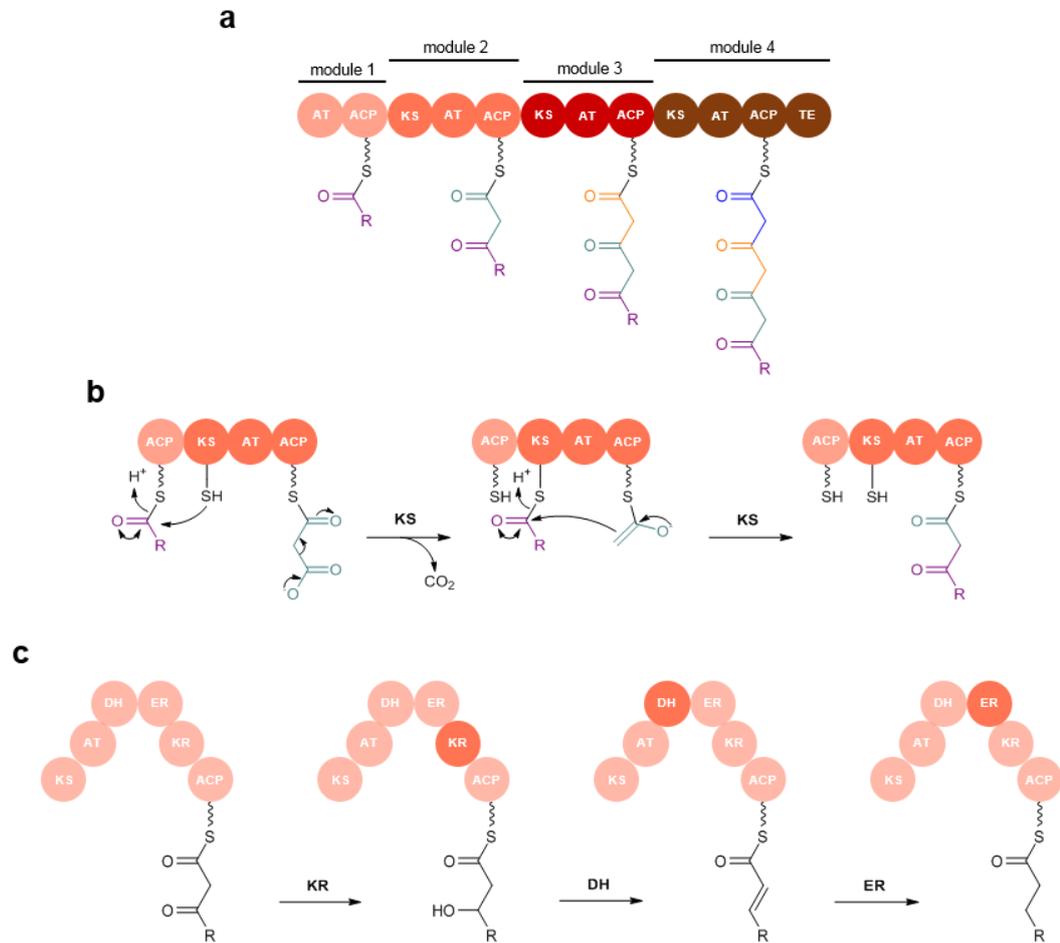


Figure 1.8. Illustration of PKS assembly line biosynthetic logic. (a) PK chain elongation on a hypothetical assembly line **(b)** The Claisen condensation catalysed by KS domains during PKS biosynthesis. Decarboxylation of the downstream (methyl)malonyl-S-ACP yields a nucleophilic thioester enolate, which attacks the upstream acyl-S-ACP thioester to form a C-C bond. **(c)** β -Carbon processing in PKSs. The β -ketoacyl-S-ACP condensation product is reduced by the KR domain to form a β -hydroxyacyl-S-ACP, subsequent dehydration by the DH domain yields an β -enoyl-S-ACP, and reduction of this species by the ER domain gives the fully saturated acyl-S-ACP.

closely related to chalcone and stilbene synthases in plants. The latter comprise a single KS domain with an active site cysteine, responsible for performing all the functions of the essential domains of type I and II PKS systems (Shimizu *et al.*, 2017).

One of the best studied *Pseudomonas* PKS is mupirocin (pseudomonic acid A), which is marketed as a topical antibiotic (Bactroban) that is active against methicillin-resistant *Staphylococcus aureus* (MRSA), and is an excellent example of how chemical diversity is generated in pseudomonad PK NPs. The pseudomonic acids A-D are a mixture of antimicrobial compounds, produced by *P. fluorescens* NCIMB 10586, that all possess a dehydroxylated tetrahydropyran ring bearing functionalised side chains in positions C-2 and C-5 (Figure 1.9). The BGC was first identified by Thomas *et al.* in 2003 (El-Sayed *et al.*, 2003) and is one of the earliest examples of a *trans*-AT PKS system (Helfrich and Piel, 2016). These PKS modules lack AT domains, but genes encoding AT(s) are present elsewhere in the cluster and act iteratively *in trans* to supply extender units to the assembly line (Figure 1.9). *trans*-AT PKS exhibit greater structural diversity than *cis*-AT systems as discrete AT domains relieve restrictions imposed by colinearity, facilitating a greater propensity for evolution in a mosaic-like fashion by extensive HGT and recombination. This is exemplified by the pederin family of compounds (Piel, 2002; Piel *et al.*, 2004; Fisch *et al.*, 2008), and the products of the misakinolide, luminaolide and tolytoxin *trans*-AT PKS systems (Ueoka *et al.*, 2015).

Structural diversification in mupirocin biosynthesis is achieved via multiple during- and post-PKS tailoring reactions, required to install the carbon at C-15, and to generate epoxide and tetrahydropyran groups. The C-15 methyl group is derived from acetate and is installed by the hydroxymethylglutaryl-CoA synthase MupH of the 'MCS cassette' (Wu *et al.*, 2007), and the epoxide is the product of the KS/oxidoreductase mmpE (Gao *et al.*, 2014). Biosynthesis of the tetrahydropyran ring remains to be completely dissected but candidate enzymes have been identified in MupC (NADH oxygenase), MupO (cytochrome P450), MupT (ferredoxin dioxygenase) and MupW (dioxygenase), with MupW representing the most likely candidate based on mutagenetic experiments (Cooper *et al.*, 2005). Finally, the mupirocin biosynthetic pathway also exhibits the use of a non-canonical extender unit 3-hydroxypropionate, which, in addition to three malonyl-CoA units, is incorporated by the iterative FAS MmpB into the 9-hydroxy-nonanoic acid moiety (Hothersall *et al.*, 2007).

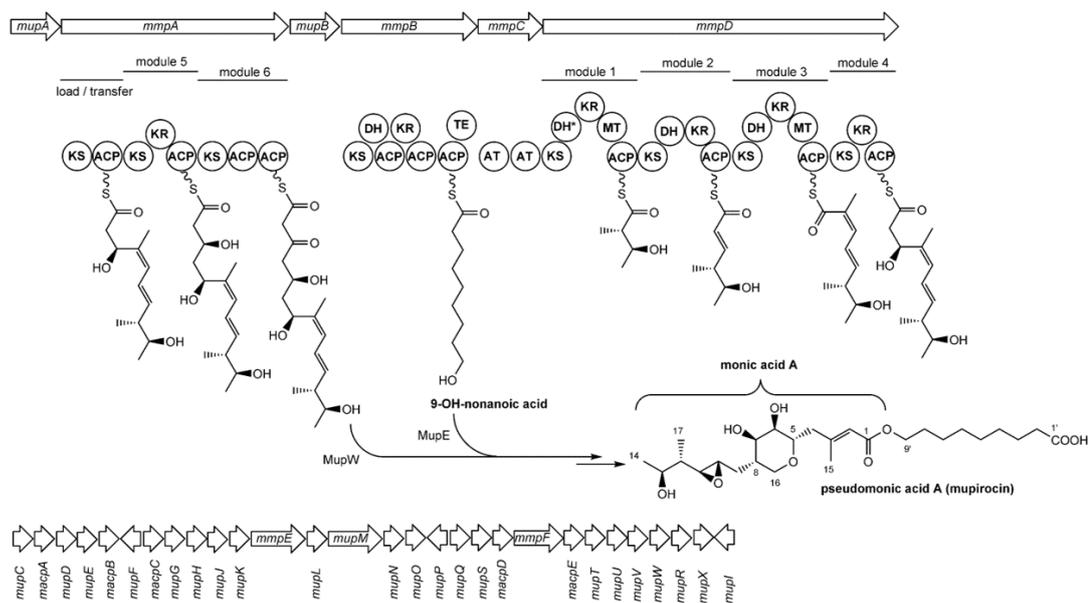


Figure 1.9. Biosynthesis model and BGC for mupirocin. * = inactive domain.
 Gross and Loper (2009) - Reproduced with permission from The Royal Society of Chemistry.

2,4-DAPG (Figure 1.4) is also of PK origin, but is synthesised by a type III PKS (Bangera and Thomashow, 1999). The key enzyme for biosynthesis is PhID which condenses three malonyl-CoA units to form activated 3,5-diketoheptanedioate. This is subsequently decarboxylated and cyclised in a Claisen condensation to yield phloroglucinol (Achkar *et al.*, 2005; Zha *et al.*, 2006), before PhiABC then install two acetyl groups at the C-2 and C-4 positions (Shanahan *et al.*, 1993). Despite their structural resemblance to phloroglucinols, the 2,5-dialkylresorcinols are not assembled on a PKS, and are instead generated by multiple modifications of medium-chain-length fatty acids derived from fatty acid metabolism (Nowak-Thompson *et al.*, 2003). These examples serve to highlight the multiple mechanisms by which building blocks derived from fatty acid biosynthesis can be assembled into NPs.

1.2.3.5 Hybrid NRP-PK NPs

The highly modular nature and analogous biosynthetic logic of NRPS and PKS systems, particularly their shared use of phosphopantetheinylated carrier protein domains to tether substrates, has allowed for the ready hybridisation of these systems to generate new NPs and further structural diversity. NRPS/PKS systems can roughly be divided into two classes: those in which an NRPS-bound peptidyl intermediate is elongated by a PKS module and vice versa (Du *et al.*, 2001). In pyoluteorin biosynthesis, L-proline is the starting unit, and this amino acid is extended by addition of three malonyl-CoA units on a PKS. L-proline is introduced by discrete A and PCP domains which activate and tether it respectively, and is the source of the dichloropyrrolyl moiety to which the three malonyl-CoA units are subsequently added (Nowak-Thompson *et al.*, 1997; 1999).

The cyclic lipopeptides (CLPs) are a structurally diverse class of compounds that minimally comprise a fatty acyl residue, ranging from five to sixteen carbons in length, attached to a peptide chain of seven to twenty-five amino acids. Between four and fourteen of these undergo macrocyclisation to form a macrolactam ring, as illustrated by massetolide A and syringomycin in Figure 1.4. (Raaijmakers *et al.*, 2006; Gross and Loper, 2009; Nguyen *et al.*, 2016). Where the first module in pyoluteorin contains a KS domain for C-C bond formation between malonyl-CoA and proline to yield an amino-acyl-S-ACP starter unit for the PKS PItB, the converse occurs in CLP biosynthesis. The first module contains a C domain for C-N bond formation between an amino acid and fatty acyl residue to yield an acylamino-S-PCP. Unusual features common to the biosynthesis of all CLPs include the presence of dual C/E domains for conversion of L- to D- configured amino acids (Balibar *et al.*, 2005), and tandem

terminal TE domains that fulfil product release, cyclisation and potentially proof-reading functions (Scholz-Schroeder *et al.*, 2003). Many CLPs occur as a mixture of analogues that differ in fatty acid and amino acid composition, again reflecting the relaxed substrate specificity of A and C domains in their biosynthetic pathways (Gross and Loper, 2009).

1.2.3.6 Shikimic acid-derived NPs

In addition to acting as noncanonical substrates to NRPS assembly lines, intermediates and derivatives of the shikimic acid pathway are themselves modified by biosynthetic enzymes to form important pseudomonad NPs. The phenazines, for example, are derived from chorismate, the major branch point in the pathways to aromatic amino acids (Knaggs, 2003). This precursor undergoes a series of modifications catalysed by products of the conserved *phzA/BDEFG* operon (Blankenfeldt *et al.*, 2004; Parsons *et al.*, 2004; Mavrodi *et al.*, 2006; Ahuja *et al.*, 2008; Guttenberger *et al.*, 2017). Chemical diversity is introduced among the phenazines through the activity of further genes either associated with the core operon, such as the *S*-adenosylmethionine (SAM)-dependent *N*-methyltransferase PhzM (Parsons *et al.*, 2007), or which are located externally like PhzH, which converts carboxylic acids to carboxamides (Mavrodi *et al.*, 2001).

The quinolones are also derived from chorismate, and are small antibacterial alkaloids that have been identified as important signalling molecules; 2-heptyl-3-hydroxy-4-quinolone was renamed *Pseudomonas* quinolone signal (PQS) due to its diverse regulon which includes genes encoding elastase, rhamnolipids, PA-IL lectin and pyocyanin, which play various important roles in biofilm formation and cellular fitness (Calfee *et al.*, 2001; Bredenbruch *et al.*, 2005; Gross and Loper, 2009). The biosynthetic locus comprises seven structural genes, *pqsABDH* and *phnAB*, that catalyse the formation of PQS from head-to-head condensation of anthranilate and β -keto-(do)decanoic acids to create a functional hybrid molecule (Gallagher *et al.*, 2002; Coleman *et al.*, 2008; Schertzer *et al.*, 2010; Drees and Fetzner, 2015; Drees *et al.*, 2016). PhnA and PhnB are thought to synthesise anthranilate from chorismate, though in some *P. aeruginosa* strains, it can be generated by L-tryptophan degradation via the kynurenine pathway (Farrow and Pesci, 2007).

1.2.4 Summary

Pseudomonas spp. exhibit an extraordinary repertoire of structurally diverse NPs (Figure 1.4) that have allowed them to colonise a wide range of environments and

adapt to a variety of different lifestyles. The advancement of modern genomics has highlighted the importance of MGEs in driving the diversification of *Pseudomonas* NPs, mediated by HGT. Evidence of this process is apparent in the mosaic arrangement of core and flexible regions of pseudomonad genomes. It is also reflected in the distribution and conservation of BGCs like that of pyrrolnitrin, which is shared among different *Pseudomonas* spp., as well as other closely related proteobacteria (Hammer *et al.*, 1999; Costa *et al.*, 2009; Weissman and Müller, 2010). Chemical diversity is achieved in the variation of number and structural composition of domains in modular biosynthetic machineries and in their inherent plasticity, underlined by unusual domain organisations and the relaxed substrate specificities observed in many of these megaenzymes. Further sophistication and divergence from the model NRPS and PKS systems in Gram-negative bacteria is achieved through the functional combination of modular and dissociated PKS and NRPS machineries. The common occurrence of *trans*-AT PKS systems in this genus (Piel, 2002; Brendel *et al.*, 2007), which exhibit greater propensity for pathway rearrangement and product diversity than *cis*-PKS systems, is conspicuous. So too is the number of pathways comprising biosynthetic enzymes for the modification of primary metabolic products, particularly amino acids and shikimic acid-derived metabolites.

Substantial changes in NP product structure can be achieved through the introduction of noncanonical building blocks, a common occurrence in pseudomonad NP biosynthesis. This can involve either the modification of canonical substrates during- and post-modular enzyme assembly, or the generation of noncanonical substrates *de novo* prior to incorporation into a larger molecule. This project concerns the antibiotic obafluorin produced by *P. fluorescens* (Figure 1.4 - Wells *et al.*, 1984; Tymiak *et al.*, 1985), which likely represents an example of the latter in which an unusual nonproteinogenic β -hydroxy- α -amino acid (β -OH- α -AA) precursor is incorporated into the final molecule (Herbert and Knaggs, 1988; 1990; 1992a and b). This moiety is significant because it ultimately gives rise to a β -lactone ring, an infrequent but highly bioactive NP pharmacophore. The final section of this Chapter introduces obafluorin and its key structural features, in addition to describing previous experimental work that forms the foundations of this project.

From this point forward, a compound numbering system will be employed that applies to all text and figures. Any number highlighted in **bold** is in reference to the **Compound Key** (page vi).

1.3 Obafluorin

1.3.1 Discovery and characterisation

In the course of screening for new β -lactam antibiotics, the metabolite obafluorin (**1**) was isolated from cultures of a plant-associated *P. fluorescens* strain in Princeton, New Jersey in 1984 (Wells *et al.*, 1984; Tymiak *et al.*, 1985). It comprises a β -lactone ring, a highly bioactive moiety, and a 4-nitrobenzyl group, both of which are relatively uncommon features in NPs. The β -lactone ring of **1** appears to be derived from a from a very unusual nonproteinogenic β -OH- α -AA, (2*S*,3*R*)-2-amino-3-hydroxy-4-(4'-nitrophenyl)butanoate (AHNB (**2**) – highlighted in Figure 1.10), and the molecule also comprises a 2,3-dihydroxybenzamide moiety, a common constituent of bacterial siderophore NPs (May *et al.*, 2001; Crosa and Walsh, 2002).

Subsequent screening for bacteria from many different locations worldwide revealed the widespread occurrence of this molecule among *Pseudomonas* strains, which may indicate high selection for its production and dispersal by HGT. However, early bioassay experiments revealed only weak antibiotic activity against a range of bacteria using traditional disc diffusion assays and complementary results were obtained in MIC testing (Wells *et al.*, 1984; Tymiak *et al.*, 1985). This is likely due to the reactivity of the molecule; **1** was found to bind to paper and is readily hydrolysed in methanolic or basic solutions, which may have confounded these early experiments by limiting diffusion and in not accurately reflecting the amounts assayed. Despite its inherent lability, **1** was shown to protect mice challenged with *Streptococcus pyogenes* (ED₅₀ = 50 mg/kg by systemic administration), and caused cell elongation of *E. coli* cells treated with sub-lethal doses. Taken together, these experimental observations indicate that **1** operates in a specific manner, rather than as a general acylating agent, in keeping with specific activities exhibited by other β -lactone NPs (Kluge and Petter, 2010; De Pascale, 2011).

This aim of this project is to dissect the biosynthetic pathway to **1** at the molecular level, with a particular focus on characterising the formation of two key structural features. The first is the unusual β -OH- α -AA precursor **2**, and the second is the β -lactone ring, which is derived from **2**, and which is associated with NPs that are known

to be potent and specific inhibitors of many important protein targets. The subsequent sections of this Chapter develop our interests in these two moieties further.

1.3.2 β -OH- α -AAs

Despite the great structural diversity of biologically important NP scaffolds, several 'privileged' structural motifs occur within them with high frequency. β -OH- α -AAs represent one of these, and are constituents and intermediates in the biosynthesis of many agriculturally and medicinally valuable bioactive NPs, as illustrated in Figure 1.10. These building blocks are considered privileged because they often comprise more than one chiral centre on which a wealth of chemical diversity can be built, and because hydroxyl groups can serve as handles for further chemical modifications that introduce further beneficial properties or bioactivities (Süssmuth and Mainz, 2017).

1.3.2.1 Synthesis

β -OH- α -AAs are thus extremely useful units for synthetic purposes and L-*threo*-3,4-dihydroxyphenylserine is itself a (pro)drug (Droxidopa) used in the treatment of Parkinson's disease (Goldstein, 2006). Although many synthetic methods have been reported for their synthesis – including methods based on enolates (Kobayashi *et al.*, 2004), glycinamides (Seiple *et al.*, 2014), glycine Schiff's base (Trost and Miege, 2014), and the aminohydroxylation of olefins (Liu *et al.*, 2013) amongst others – these elegant methodologies often suffer from insufficient stereoselectivity, expensive precursors and/or the need for stoichiometric amounts of chiral auxiliaries (Steinreiber *et al.*, 2007a). The production of β -OH- α -AAs by enzymatic means is thus highly desirable as it allows two stereocentres to be set in one single reaction, that can be performed in a one-pot process, with minimal protection of substrates and under mild aqueous conditions.

Consequently, considerable effort has been put into developing L- and D-threonine aldolases (TAs) as biosynthetic tools for the generation of β -OH- α -AAs. TAs use the cofactor pyridoxal-5'-phosphate (PLP) to catalyse the retro-aldol cleavage of threonine to generate glycine and acetaldehyde. These enzymes supplement the role of serine hydroxymethyltransferase (SHMT), a key enzyme in the generation of C₁ units (formaldehyde) for tetrahydrofolate (THF)-dependent reactions, which cleave L-serine via an analogous mechanism (Figure 1.11 - Schirch *et al.*, 1985). Due to the reversibility of the retro-aldol reaction, TAs have been studied as potential biocatalysts for the asymmetric synthesis of L- and D- β -OH- α -AAs from achiral

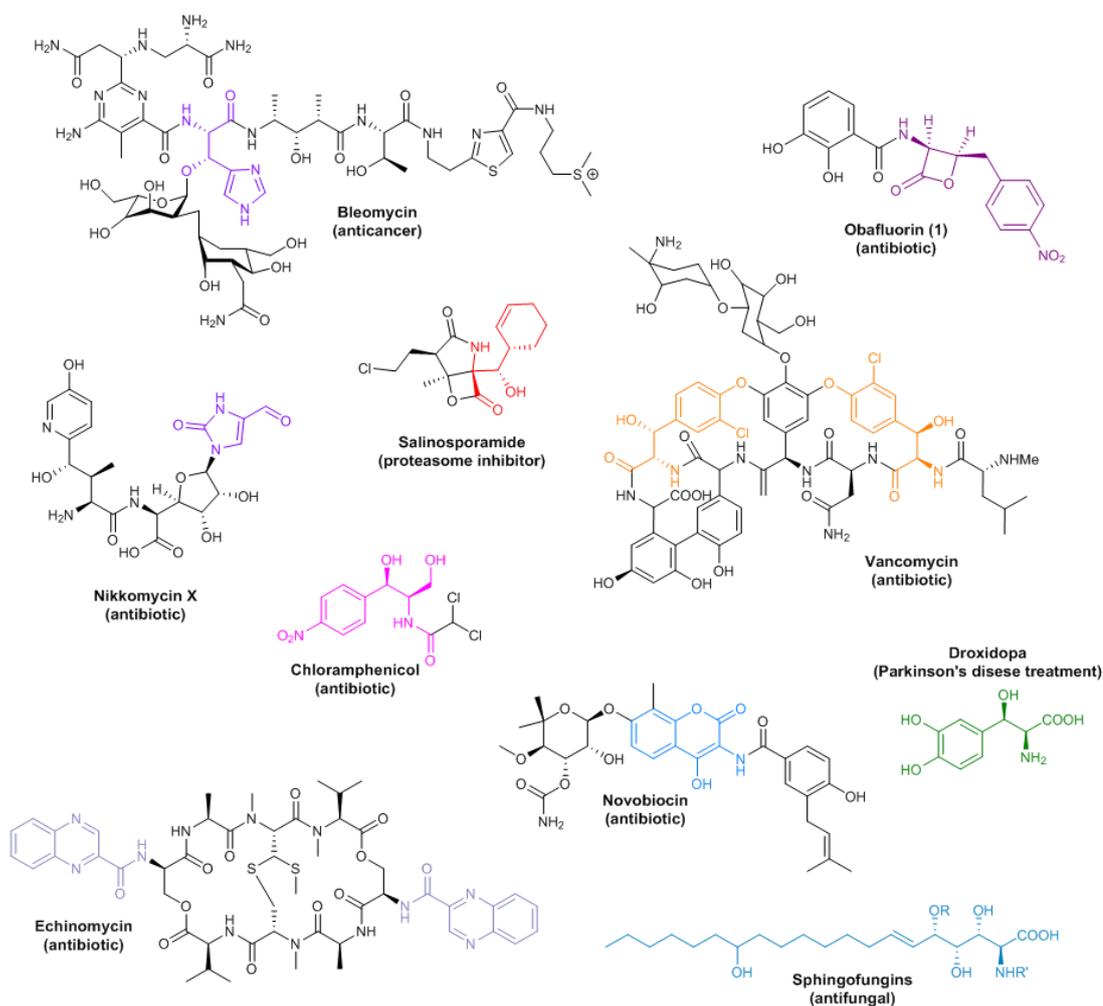


Figure 1.10. Examples of β -OH- α -AA-comprising NPs. Moieties derived from β -OH- α -AA moieties are highlighted in colour.

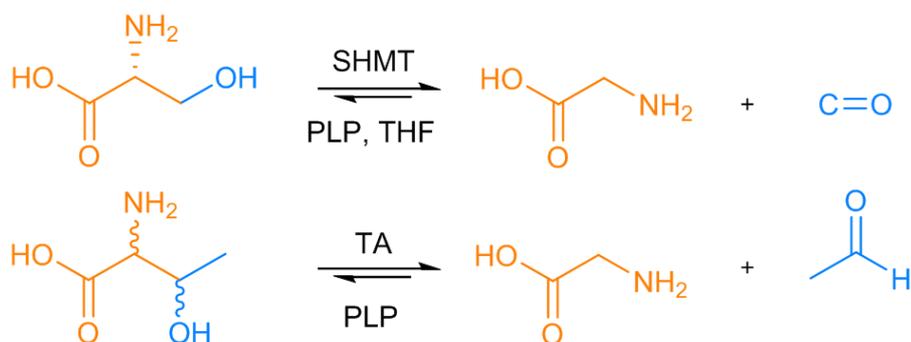


Figure 1.11. Reactions catalysed by SHMTs and TAs.

aldehydes and glycine (Dückers *et al.*, 2010; Fesko and Gruber-Khadjawi, 2013; Fesko, 2016). Whilst very selective in terms of their donor glycine substrate, most TAs will accept a range of aliphatic and aromatic aldehydes as acceptor substrates, making them promising versatile biocatalysts (Fesko *et al.*, 2008). However, despite great progress in the development of these enzymes, their utility remains limited due to only modest diastereoselectivity (highly selective for the α -carbon) and low synthetic yields (Fesko, 2016). An NP biosynthetic enzyme that has evolved to generate these building blocks efficiently with high stereoselectivity would thus represent an ideal alternative to these inefficient aldolases.

1.3.2.2 β -OH- α -AAs in NP biosynthesis

In NP biosynthesis, many β -OH- α -AAs are formed via the β -hydroxylation of proteinogenic L-amino acids. The glycopeptide antibiotic vancomycin, for example, comprises two β -hydroxytyrosine units in its peptide backbone (the 2*R*,3*R*-3-OH diastereomer at the C-2 position, and the 2*S*,3*R*-3-OH isomer at position C-6 – Figure 1.10). β -Hydroxylation of tyrosyl-S-ACP is catalysed by the cytochrome P450 OxyD, following the activation and tethering of L-tyrosine to the A-PCP didomain protein VcmD. The use of aminoacyl-S-PCP intermediates as the substrates for P450 monooxygenase-catalysed β -hydroxylation is a common mechanism for β -OH- α -AA generation in NPs. Further examples include the formation of β -hydroxytyrosine in aminocoumarin-containing angucycline antibiotics such as novobiocin (Chen and Walsh, 2001), β -hydroxyhistidine in nikkomycin biosynthesis (Chen *et al.*, 2002), and β -hydroxytryptophan in echinomycin biosynthesis (Zhang *et al.*, 2013) (Figure 1.10). In chloramphenicol, a nonheme diiron monooxygenase CmlA performs the role, catalysing the β -hydroxylation of its PCP-bound substrate 4-aminophenylalanine (4-APA; **3**) (Makris *et al.*, 2010).

Another mechanism of β -OH- α -AA biosynthesis in the context of specialised metabolism that has been reported is the reaction catalysed by L-threonine transaldolases (L-TTA). These enzymes are extremely rare and there are currently only two characterised examples: fluorothreonine transaldolase (FTase) and members of the LipK family of L-threonine:uridine-5'-aldehyde transaldolases (Murphy *et al.*, 2001; Barnard-Britson *et al.*, 2012). Like L-TAs these are PLP-dependent enzymes that catalyse the cleavage of L-threonine to generate glycine and acetaldehyde (Figure 1.10). However, they are unique in that they subsequently catalyse C-C bond formation between the enzyme-bound glycine unit and an aldehyde substrate to yield β -OH- α -AA products. This results in the generation of

L-fluorothreonine in the case of FTase and an unusual 5'-C-glycyuridine (GlyU) unit for LipK and its homologues.

Based on the results of isotope labelling studies described later in this introduction, it is unlikely that the assembly of **2** during the biosynthesis **1** is the result of a cytochrome P450-catalysed oxidation. Instead a thiamine diphosphate (ThDP)-dependent mechanism has been proposed (Herbert and Knaggs, 1990) that is described in section 1.3.4 and would also preclude an L-TTA-mediated mechanism. However, these enzymes are revisited and discussed in further detail in Chapters 5 and 6. The pathway to **2** thus potentially represents a completely new mechanism of β -OH- α -AA biosynthesis.

1.3.2.3 Project outlook

Identifying the biosynthetic mechanism by which **2** is generated is of interest for two key reasons. The first is that an enzyme involved in the biosynthesis of **2** may represent a valuable query for genome mining. P450 oxygenases belong to an enormous superfamily of oxidative enzymes that play widespread tailoring functions in NP biosynthetic pathways (Podust and Sherman, 2012). They are therefore likely to be limited as queries for genome mining efforts to identify further β -OH- α -AA-containing NP BGCs due to high chances of generating false positive hits. The enzyme responsible for the biosynthesis of **2** could therefore be an ideal alternative if it represents a far more specific mechanism, and could be applied for the identification of further BGCs in which β -OH- α -AAs are similarly biosynthesised.

The second reason is that current biocatalysts for the preparation of β -OH- α -AAs are not particularly efficient or stereospecific. Given the significance of **2** in formation of the β -lactone ring of **1**, it is likely that the enzyme responsible is stereospecific to avoid the production of inactive products. It may thus represent an ideal candidate for development as a biocatalyst for both synthetic and industrial applications. If the enzyme responsible for the biosynthesis of **2** is permissive to alternate substrates, it might be possible to generate completely novel β -OH- α -AA building blocks and provide access to brand new chemical space. Depending on the mechanism, characterisation of such an enzyme might also create opportunities to search for, or synthetically produce, novel β -lactone metabolites.

1.3.3 β -Lactones

The β -lactone structural motif occurs infrequently in nature but is associated with molecules possessing potent and valuable biological activities (Figure 1.12) (Kluge and Petter, 2010). They are highly electrophilic ring systems, readily susceptible to attack by nucleophilic protein amino acid residues (Figure 1.13). This makes them powerful acylating agents and they exhibit enormous therapeutic potential as highly effective, specific inhibitors of many important enzyme targets (De Pascale *et al.*, 2011). Given the impressive bioactivity of these ring-strained systems, they have arisen from multiple biosynthetic origins in nature. Subsequent sections cover several of the better genetically and biochemically characterised examples of these molecules, describing their biological targets and diverse biosynthetic pathways, with a focus on the variety of proposed mechanisms of β -lactone formation.

1.3.3.1 Ebelactones

The ebelactones are produced by *Streptomyces aburaviensis* and have been shown to be inhibitors of several important protein targets including esterases (Umezawa *et al.*, 1980), lipases (Nonaka *et al.*, 1996), cutinases (Koller *et al.*, 1990), homoserine transacetylase (De Pascale *et al.*, 2011) and cathepsin A (Ostrowska *et al.*, 2005). In the case of homoserine transacetylase, inhibition is mediated by covalent modification of an active site serine residue (De Pascale *et al.*, 2011).

This surprisingly broad range of enzyme targets triggered interest in this group of compounds and their biosynthesis, and the BGC was finally identified in 2013. The compound is assembled across a PKS composed of six modules, each with a KS domain for chain elongation. KR, DH and ER domains present are almost all correctly placed to perform the necessary processing of the growing chain to introduce β -hydroxy, alkene and methylene groups (Wyatt *et al.*, 2013 - Figure 1.14). Module 1 comprises a characteristic KS^Q domain which loads malonyl-CoA before decarboxylating the malonyl-enzyme product to yield an acetate starter unit (Bisang *et al.*, 1999). All other AT domains have specificity for methylmalonyl-CoA, again consistent with the final structure. One noncanonical element of ebelactone biosynthesis is in the organisation and use of module 4. Its cognate domains are split across three genes *ebeDEF* and it is anticipated to be used twice. In the first round, only the KS is used, leaving the required ketone intact. In the second iteration, the KS is used again but the growing chain is now processed by the discrete DH EbeE and the N-terminal KR of EbeF, which comprises the rest of modules 4 and 5.

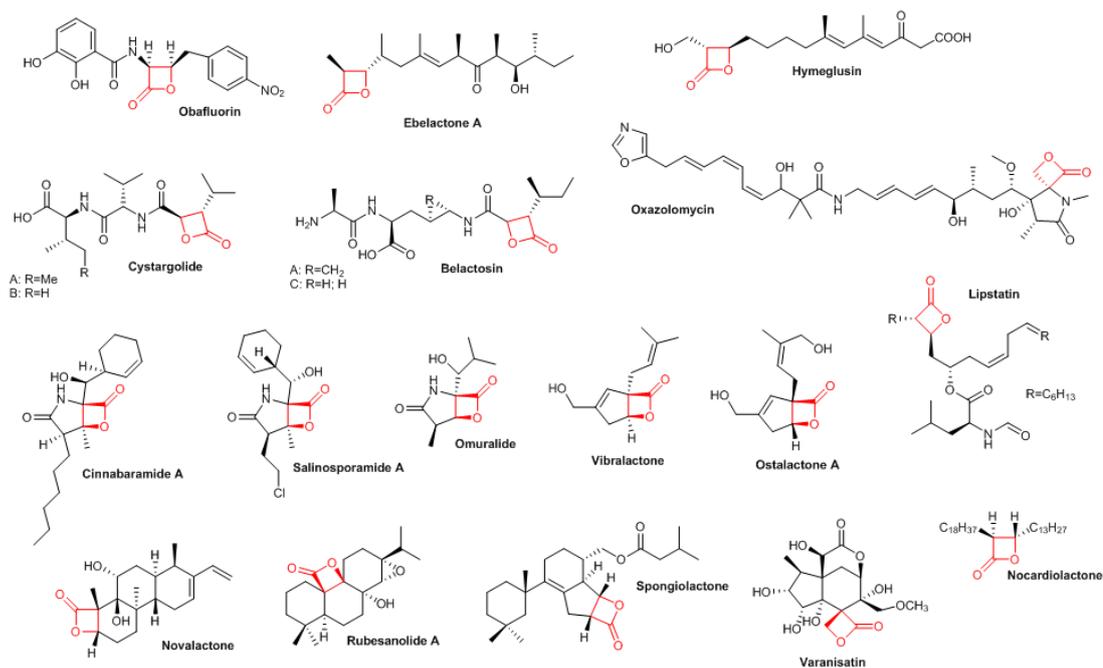


Figure 1.12. Examples of β -lactone NPs. β -Lactone moieties are highlighted in red.

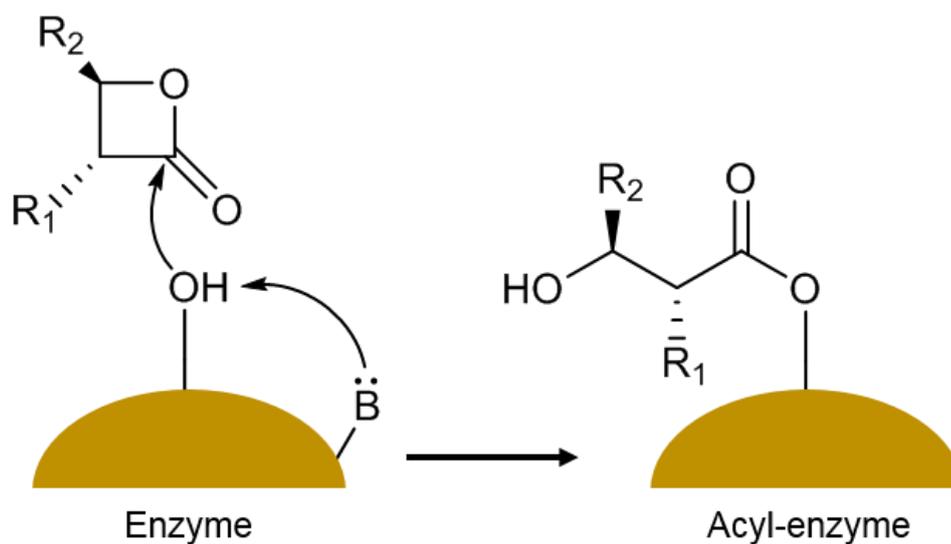


Figure 1.13. General mechanism of β -lactone acylation. An active site nucleophile (serine represented) attacks the β -lactone ring ester to generate a covalently bound acyl-ester, preventing normal enzyme function.

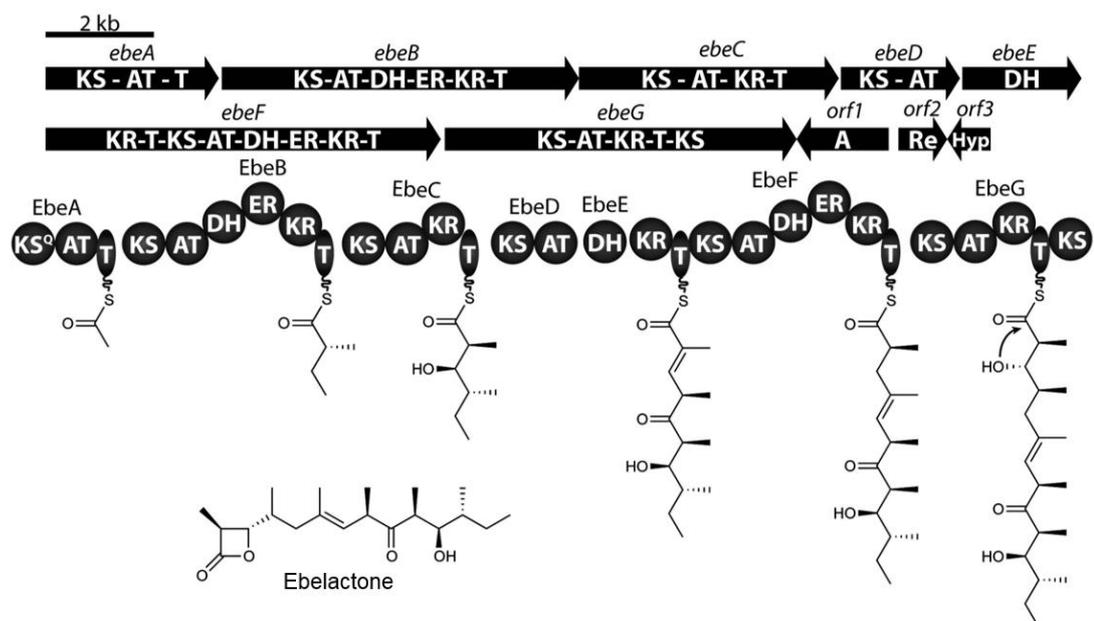


Figure 1.14. Biosynthetic model and BGC for ebelactone. Shown are the putative genes for ebelactone biosynthesis and the corresponding PKS assembly line. Downstream genes of possible biosynthetic relevance are shown although their putative functions are not required in the final maturation of ebelactone. These include an A domain, a reductase (Re), and a hypothetical protein (Hyp). Tethered PK intermediates are shown on the ebelactone assembly-line. The proposed ebelactone release mechanism through β -hydroxy attack is illustrated. T = thiolation (ACP) domain. Wyatt et al. (2013) - Reproduced with permission from the Nature Publishing Group.

Intriguingly, there is no terminal TE domain to catalyse β -lactone ring formation and product release (Figure 1.14), and the spontaneous cyclisation of the final β -ketoacyl-S-ACP intermediate was proposed, though this reaction would be energetically disfavoured. Alternatively, a terminal KS domain could potentially perform a catalytic role, but was suggested instead to be an artefact of the 'original' NP PKS in which the remaining modules have been lost following derailment by chemical cyclisation (Wyatt *et al.*, 2013). Recently however, the β -lactone synthetase OleC involved in olefin biosynthesis, was shown biochemically to catalyse the cyclisation of *syn* and *anti*- β -hydroxy acid substrates to yield *cis*- and *trans*- β -lactone olefin intermediates (Christenson *et al.*, 2017- Figure 1.15a). An *oleC* homologue (46% identity), *orf1*, was identified immediately adjacent to the ebelactone BGC (Figure 1.15b) and currently represents the most likely candidate for β -lactone ring formation.

1.3.3.2 Lipstatin

Lipstatin, first isolated from *Streptomyces toxitricini* (Weibel *et al.*, 1987), is a highly potent irreversible inhibitor of human pancreatic lipase. It comprises a 2,3-*trans*-disubstituted β -propiolactone with two linear alkyl chains (α -site C-6, β -site C-13). Tetrahydrolipstatin (Orlistat), a saturated derivative of lipstatin, inhibits mammalian lipases via covalent binding to an active site serine residue (Hadvary *et al.*, 1991). The β -lactone ring is critical for enzymatic inhibition, as opening it leads to the complete loss of lipase inhibitory activity (Stadler *et al.*, 1992). Today, it remains the only currently available FDA-approved oral drug (Xenical) for long-term treatment of obesity, mediated via control of diabetes (Torgerson *et al.*, 2004; Kang and Park, 2012). It also exhibits antitumor activity via inhibition of FAS TE domains in tumour cells (Kridel *et al.*, 2004). Other natural homologues include the panclicins (Yoshinari *et al.*, 1994), valilactone (Kitahara *et al.*, 1987) and esterastin (Umezawa *et al.*, 1978; Kondo *et al.*, 1978), which differ only in the structure of their side chains and the nature of their linked amino acids.

In silico analysis of the putative lipstatin BGC combined with the results of stable isotope and deuterium labelling experiments allowed a hypothetical pathway to be proposed (Eisenreich *et al.*, 1997; 2003; Goese *et al.*, 2000; 2001; Schuhr *et al.*, 2002; Bai *et al.*, 2014 – Figure 1.16). In the first committed step, Claisen condensation between 3-hydroxytetradeca-5,8-dienoyl-CoA and hexyl-malonyl-CoA, both derived from linolenic acid, yields the C₂₂ α -branched fatty acid backbone. Subsequent carboxylation by an unidentified acyl-CoA carboxylase, generates hexyl-malonyl-CoA, which is subsequently proposed to be transferred to a primary metabolic ACP.

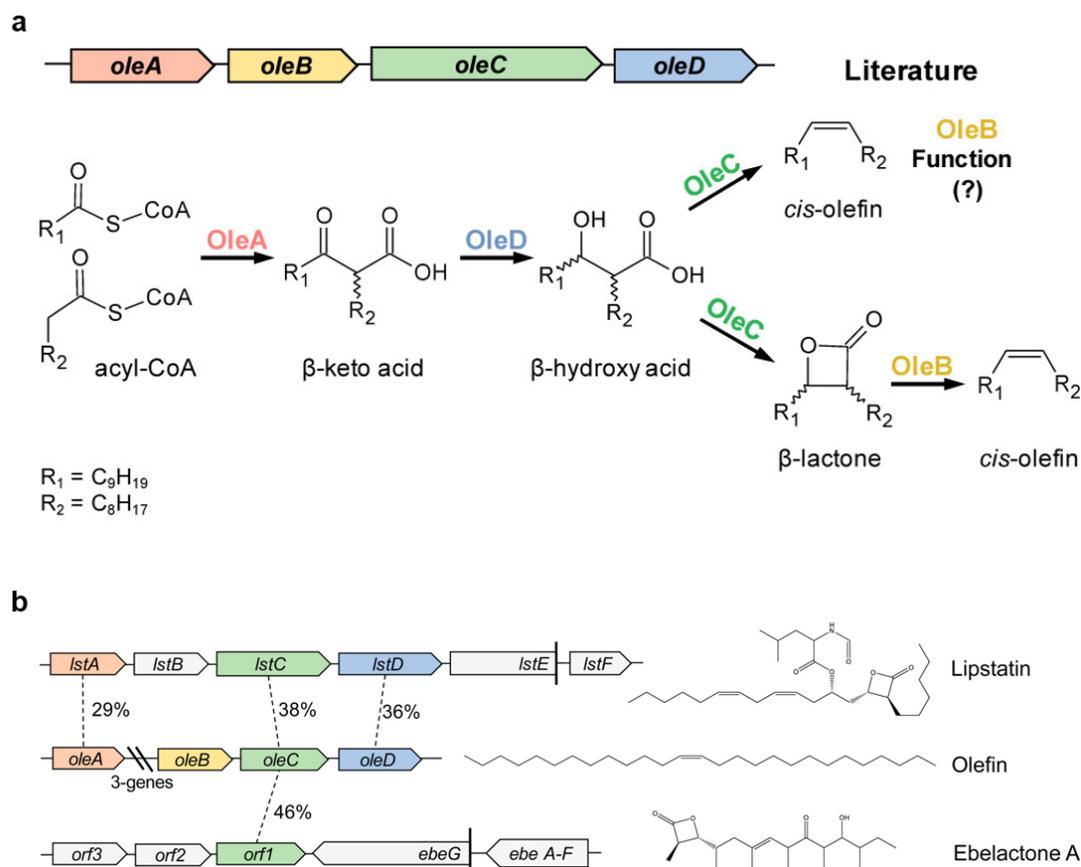


Figure 1.15. Illustration of β -lactone ring formation catalysed by OleC. (a) Illustration of the ole BGC and enzymatic pathway. (b) Homologous OleC enzymes encoded in the ebelactone and lipstatin BGCs. Percent identity is based on amino acid sequences. Adapted with permission from Christenson et al. (2017) *Biochemistry* 56, 348-351. Copyright (2017) American Chemical Society.

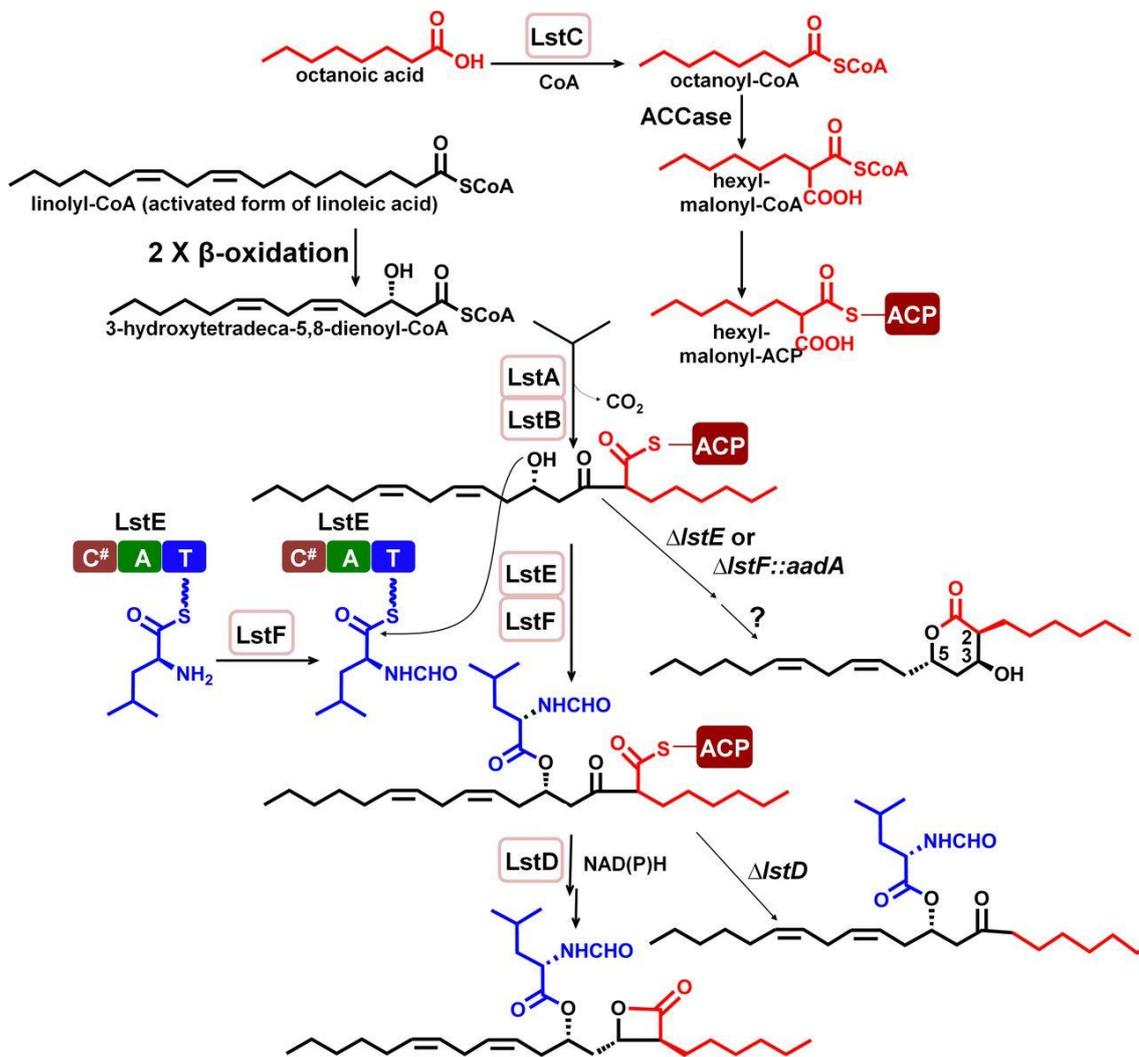


Figure 1.16. Biosynthetic model for lipstatin. Bai et al., (2014) - Reproduced with permission from the American Society for Microbiology.

LstAB, β -ketoacyl-ACP synthase III FabH homologues, are proposed to subsequently decarboxylate hexyl-malonyl-CoA and catalyse condensation with 3-hydroxytetradeca-5,8-dienoyl-ACP to afford a 3-keto-5-hydroxy-C₂₂-ACP intermediate. A monomodular NRPS LstE activates and tethers L-leucine which is formylated by the formyltransferase LstF. The LstE C domain is then thought to catalyse nucleophilic attack of the 5-hydroxyl group of the 3-keto-5-hydroxy-C₂₂-ACP intermediate on the acyl carbon of formyl leucine-S-PCP, by an analogous mechanism to AT reactions.

The β -lactone ring was initially proposed to form spontaneously by attack of the 3-hydroxy group on the carbonyl of the resulting ACP-tethered intermediate, following LstD (a 3-hydroxysteroid dehydrogenase/isomerase homologue)-catalysed reduction of the 3-keto group. However, like ebelactone, an *oleC* homologue (*IstC*) was identified in the *Ist* BGC (Christenson *et al.*, 2017 – Figure 1.15b), indicating an energetically more favourable catalytic route to β -lactone ring formation, but in a completely different BGC context.

1.3.3.3 Salinosporamides and analogues

The γ -lactam- β -lactone NP family are highly potent inhibitors of the 20S proteasome, the key proteolytic component of the ubiquitin-mediated protein degradation pathway (Groll *et al.*, 2005). They include the salinosporamides, cinnabaramides and omuralide, which all share an *ortho*-fused γ -lactam- β -lactone bicyclic ring system core structure, but differ in the groups attached at the C-2, C-3 and C-5 positions (Figure 1.12). They all undergo the same initial reaction step with their target in which the N-terminal threonine residue of the 20S proteasome catalytic β_5 -subunit attacks the β -lactone ring, resulting in the formation of a covalent ester linkage (Groll *et al.*, 2006a). Binding is made irreversible in the case of salinosporamide A by a subsequent reaction in which the β -lactone-derived C-3 hydroxyl group reacts with a chlorinated C-2 side chain leaving group. This yields a tetrahydrofuran adduct that is stable to water hydrolysis. Omuralide is only a slow reversible inhibitor, in which the 5-hydroxy group of the opened β -lactone ring acts to displace water from the active site and prevent it from re-entering (Shah *et al.*, 2002). This highlights the importance of the rest of the molecule in determining β -lactone inhibitory specificity and efficacy. Salinosporamide A (Marizomib) thus has enhanced potency (IC₅₀ value of 1.3 nM vs 49 nM for omuralide) and is currently in phase I human clinical trials for the treatment of multiple myeloma (Feling *et al.*, 2003; Harrison *et al.*, 2016). Structurally similar NPs include vibractone (Liu *et al.*, 2006) and ostalactone (Kang and Kim, 2016)

from the basidiomycetes *Boreostereum vibrans* and *Stereum ostrea* respectively, though these lack the γ -lactam functionality in their 4/5 fused ring systems and appear to be more potent inhibitors of lipases.

Salinosporamide A was discovered in 2003 from the marine actinomycete *Salinospora tropica* CNB-392, which was subsequently found to produce several other salinosporamide analogues (Feling *et al.*, 2003). It is assembled by a mixed PKS/NRPS system that is focussed around two key megasynth(et)ase-encoding genes, *salAB* (Figure 1.17 - Udvary *et al.*, 2007; Eustáquio *et al.*, 2009). SalA is a dimodular PKS with an unusual domain organisation that comprises continuous loading and extender AT domains. The loading AT_L domain is of the type that recognises monocarboxylic acid starter units and is proposed to use acetyl-CoA to form an ACP-bound intermediate. Phylogenetically, the AT₁ domain is placed in the methylmalonyl-CoA accepting-group and contains a distinct signature motif that is consistent with it accepting an unreported extender unit. In salinosporamide A, this is chloroethylmalonyl-CoA which is the product of eight enzymatic steps catalysed by the products of *salGHLMNQST* (Eustáquio *et al.*, 2008). However, the detection of analogues with different side chains at C-2 points to a degree of promiscuity in this AT domain (Liu *et al.*, 2009). The SalA KS₁ domain presumably catalyses Claisen condensation between the two tethered substrates before the terminal C₂ domain catalyses peptide bond formation between their condensation product and 3-cyclohex-2'-enylalanine, which is activated by and tethered to the A₂-PCP didomain NRPS SalB (Figure 1.17).

There is no OleC homologue associated with the *sal* BGC, indicating a completely different mechanism for β -lactone ring formation in this family of NPs. The discrete TE SalF would be a candidate to catalyse ring closure, as would its homologue *cinE* in biosynthesis of cinnabaramide (using a β -hydroxyacyl-enzyme substrate - Rachid *et al.*, 2011), however it has been suggested that these TEs instead perform an editing role to remove stalled pathway intermediates. An alternate hypothesis is that a spontaneous tandem aldol-lactonisation leads to ring formation in salinosporamide A (Gulder and Moore, 2010), however no mechanism has been experimentally validated.

1.3.3.4 Oxazolomycin and analogues

Oxazolomycin A from *Streptomyces albus* JA3453 (Mori *et al.*, 1985) also comprises a fused γ -lactam- β -lactone bicyclic ring system, but differs in that the two rings are

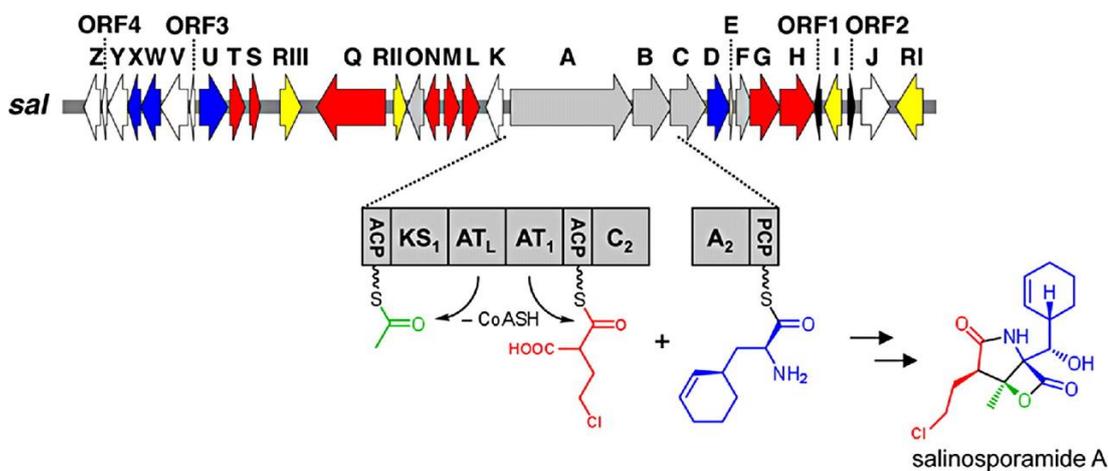


Figure 1.17. Biosynthetic model and BGC for salinosporamide A. Genes putatively involved in the chloroethylmalonyl-CoA pathway (red), construction of the core γ -lactam- β -lactone ring system (grey), assembly of the nonproteinogenic amino acid L-3-cyclohex-2'-enylalanine (blue), regulation and resistance (yellow), unknown (white), and 2 partial transposases (black) are color-coded. Eustáquio et al., (2009) – Adapted with permission from the Proceedings of the National Academy of Sciences of the USA.

spiro-linked (Figure 1.12). The structure is also characterised by a 5-substituted oxazole ring, an (*E, E*)-diene, and a (*Z, Z, E*)-triene chain. The producing strain synthesises two other geometrical isomers B and C that differ only in the configuration of the triene moiety (*E, E, E* and *Z, E, E* respectively) (Kanzaki *et al.*, 1998). Other family members include 16-methyloxazolomycins KSM-26908 and KSM20906 (Otani *et al.*, 2000), triedimycins (curryomycins) A and B (Ikeda *et al.*, 1991; Ogura *et al.*, 1985), inthomycins (phthoxazolins) A-C (Henckel and Zeeck, 1991; Shiomi *et al.*, 1995), neooxazolomycin (Takahashi *et al.*, 1985), and the lajollamycins (Manam *et al.*, 2005; Ko *et al.*, 2014). This family of compounds, all derived from various *Streptomyces* sp. strains, have been shown to exhibit diverse antibacterial, antitumor, anti-HIV and even antiviral activities (Nakamura *et al.*, 1994; Zhao *et al.*, 2006; Bagwell *et al.*, 2008).

Isotope labelled feeding experiments established that oxazolomycin is derived from three glycine and nine acetate molecules, and that all five C-methyl groups, the O-methyl group and the N-methyl group are derived from L-methionine (Zhao *et al.*, 2006). Biochemical and genetic analysis indicated that the C3-C4 unit is derived from a metabolic intermediate of the glycolytic pathway, which serves as a precursor of an unusual methoxymalonyl-ACP extender unit in the biosynthesis of several PK NPs (Wu *et al.*, 2000; Ligon *et al.*, 2002; Yu *et al.*, 2002; Rascher *et al.*, 2003; Li *et al.*, 2006). Using these experimental observations, the *ozm* BGC was identified (Zhao *et al.*, 2010 - Figure 1.18), and features *trans*-AT PKSs (OzmJKMNQ), NRPSs (OzmO and L) and a hybrid NRPS-PKS megasynthase (OzmH). Biosynthesis is proposed to be initiated by OzmO, an NRPS loading module comprising canonical A and PCP domains, as well as an N-terminal formylation domain, to activate, tether and formylate glycine to generate formyl-glycyl-S-PCP. This moiety is later cyclised to form the oxazole ring. This intermediate is then hypothesised to be transferred sequentially to OzmQ, N, H, J, K and L, where condensation of five malonyl-CoAs, a glycine, three malonyl-CoAs, a methoxymalonyl-ACP, a malonyl CoA and an L-serine occurs, in addition to C- and N- methylation events.

Oxazolomycin likely represents yet another mechanism of β -lactone ring formation. The final module of OzmL contains a terminal C domain that is proposed to be responsible for concomitant formation of the *spiro*-linked γ -lactam- β -lactone moiety and release of the final product, as a TE domain is absent from this termination module (Figure 1.18). This terminal C domain lacks a full catalytic triad for peptide bond formation, but does comprise the aspartate residue that is conserved in both C

and Cy domains (Zhao *et al.*, 2010). As described previously, the latter is responsible for the cyclisation of either cysteine, serine or threonine residues to generate thiazoline or oxazoline rings. This C domain is therefore hypothesised to catalyse cyclisation of the oxazolomycin serine side chain to form the β -lactone ring.

1.3.3.5 Belactosins and cystargolides

As a final example, the belactosins and cystargolides represent a family of β -lactones which, like the *ortho*-fused γ -lactam- β -lactones, are proteasome inhibitors that mediate their effects through covalent binding to the proteasome β_5 -subunit (Asai *et al.*, 2000; Groll *et al.*, 2006b). Both consist of an *N*-acylated dipeptide and a 2-carboxy-3-alkyl β -lactone moiety, but slightly different dipeptide backbones (Figure 1.12). However, sequencing of the cystargolide-producer *Kitasatospora cystarginea* NRRL B16505 genome by Wolf *et al.* (2017) did not reveal any NRPS BGCs of the predicted size and specificity. Feeding experiments with [$^{13}\text{C}_5$]-labelled L-valine revealed the incorporation of ten and fifteen carbons in cystargolides A and C respectively. Subsequent feeding with [1,2- $^{13}\text{C}_2$]-labelled acetate indicated the incorporation of five carbons from L-valine and two from acetate in the β -lactone moiety of both analogues, suggesting the product is the result of a condensation reaction between a valine metabolite and acetate. The BGC was finally identified based on the presence of two isopropylmalate synthases (IPMSs) present within the cluster. In L-leucine biosynthesis, these enzymes catalyse Claisen-type condensations between acetyl-CoA and α -ketoisovalerate, a degradation product of L-valine, to generate 2-isopropylmalate (2-IPM). The belactosin BGC was similarly identified in the genome sequence of its producer *Streptomyces* sp. UCK 14.

Based on a combination of *in silico* and biochemical data, analogous 4-step routes to β -lactone ring formation in both NP biosynthetic pathways were proposed (Wolf *et al.*, 2017 - Figure 1.19). First, the IPMS homologue CysA generates 2-IPM as described above – in belactosin biosynthesis BelJ, also an IPMS homologue, uses 2-keto-3-methylvalerate (an intermediate of isoleucine metabolism), instead of α -ketoisovalerate, to generate 2-(*sec*-butyl)-2-hydroxysuccinic acid. In the cystargolide pathway 2-IPM is isomerised to 3-IPM, presumably by primary metabolic enzymes, where in belactosin, this function is suggested to be provided by BelF and/or BelG (isopropylmalate dehydratase homologues). Putative SAM-dependent methyltransferases CysG and BelR are then proposed to form methyl esters of 3-IPM and 3-(1-methylpropyl)malate respectively, thereby generating intermediates to facilitate subsequent lactonisation and β -lactone ring formation. Alternatively, putative

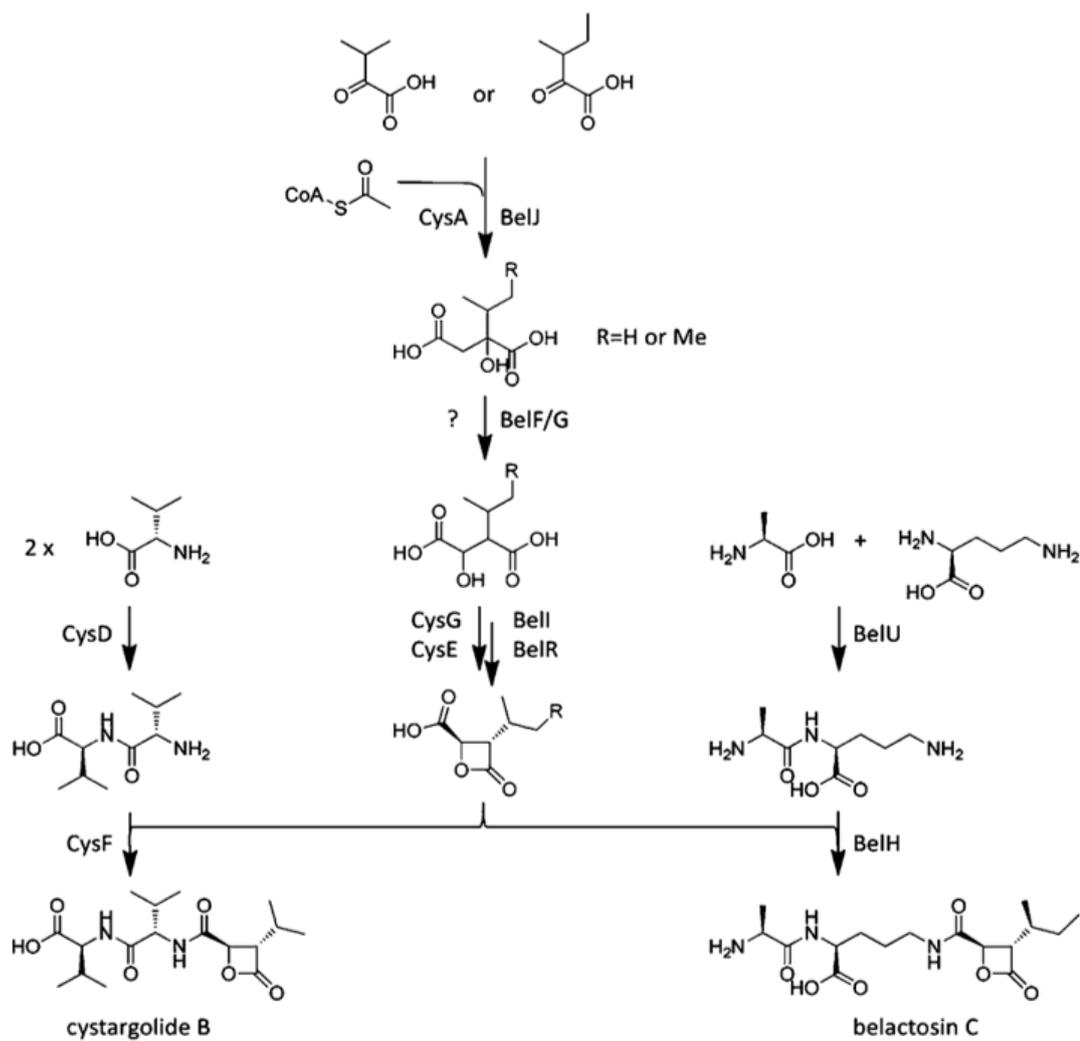


Figure 1.19. Biosynthetic models for belactosin C and cystargolide A. Wolf et al. (2017) – Reproduced with permission from John Wiley and Sons, Inc.

acyl-CoA ligases CysC and BelV could activate the appropriate carboxylic acids for lactonisation. Putative esterases CysE and BelR are then thought to be responsible for supporting intramolecular esterification and cyclisation (Wolf *et al.*, 2017). The dipeptide backbones of both molecules are hypothesised to be assembled by ATP-grasp ligases CysD and BelU. These enzymes have only recently been shown to be involved in amide bond formation in NP biosynthesis (Dawlaty *et al.*, 2010; Noike *et al.*, 2015; Ooya *et al.*, 2015), traditionally playing roles in primary metabolism such as in fatty acid, glutathione and cell wall biosynthesis (Fawaz *et al.*, 2011).

1.3.3.6 Terpenoid β -lactones

Given that β -lactones interact so readily and effectively with their protein targets, it is perhaps unsurprising that multiple biosynthetic routes to their generation have arisen in nature. Other NP examples not described here include the 3-hydroxy-3-methylglutaryl-CoA synthase inhibitor hymeglusin (Greenspan *et al.*, 1987; Skaff *et al.*, 2012) and the Hsp70 inhibitor novolactone (Hassan *et al.*, 2014), both of fungal origin, the rubesanolides (Zou *et al.*, 2011; 2012) and veranisatin (Yin *et al.*, 2012) from the plant species *Isodon rubescens* and *Illicium simonsii* respectively, and spongiolactone from the Mediterranean sponge *Spongionella gracilis* (Mayol *et al.*, 1987) (Figure 1.12). All but hymeglusin appear to be of terpenoid origin, derived from the C₅ building block isopentenyl pyrophosphate (IPP), which indicates that further mechanisms of β -lactone ring formation remain to be characterised. However, these examples are yet to be studied biosynthetically.

1.3.3.7 Potential of synthetic β -lactones

Considerable effort has also been put into generating synthetic β -lactones either to improve potency of naturally occurring compounds, to improve drug-like characteristics, or for use as building blocks for the biosynthesis of other synthetic molecules (Soucy *et al.*, 1999; Yang and Romo, 1999; Schneider, 2002). Stefan Sieber and his lab were able to employ the reactivity of the β -lactone ring in activity-based protein profiling (ABPP) efforts to identify new therapeutic targets and to gain a deeper understanding of the molecular mechanisms governing pathogenesis (Böttcher and Sieber, 2012). By generating a synthetic library of β -lactones inspired by naturally occurring compounds, and comprising an alkyne tag, they could probe the proteomes of Gram-positive and Gram-negative strains related to pathogenic strains, and identified many interesting cellular targets (Böttcher and Sieber, 2008a; 2008b). Among these was the virulence associated protease ClpP, known to control the regulation of many virulence factors including hemolysins, proteases and

DNAases, and to which their β -lactone probes were bound via an active site serine residue. Probes were shown to be potent ClpP inhibitors, leading to attenuated virulence determinant production in both *S. aureus* and *L. monocytogenes*, and could be made more potent by altering C-2 and C-3 side chain substituents (Böttcher and Sieber, 2008a; 2008b; 2009a; Gersch *et al.*, 2013). This work illustrates the enormous potential of synthetic β -lactones for the inhibition of novel biological targets, and reiterates the significance of stereochemical configuration and side chain composition in determining their specificity and potency.

1.3.3.8 Project outlook

Identifying new enzymes and mechanisms for the biosynthesis of β -lactones has important implications for synthetic biology approaches to generate novel protein inhibitors. Characterisation of the mechanisms employed during the biosynthesis of **1** may provide a novel route by which these biologically active moieties can be synthetically introduced into new metabolites. Furthermore, if unique, the mechanism might also provide a suitable query for genome mining efforts to identify further β -lactone NPs. The bioactivity conferred upon **1** by comprising such a reactive electrophilic moiety, and the influence the aromatic side chains have on governing target specificity and bioactivity, are also of interest, particularly given the number and distribution of isolated **1**-producers identified by Wells *et al.* (1984). Therefore, investigating the bioactivity and mechanism of action for **1** is also within the scope of this project. Nature has found multiple biosynthetic routes to incorporate these moieties into highly bioactive metabolites, indicating that the presence of a β -lactone ring in **1** is unlikely to be trivial. This is particularly so given that the pathway has evolved to utilise an unusual β -OH- α -AA to generate it.

1.3.4 Preliminary investigations of obafluorin (**1**) biosynthesis

The presence of such a bioactive moiety and its potentially unique biosynthetic origins prompted chemical investigations to determine the likely biosynthetic pathway to **1**. Stable isotope feeding experiments performed on the **1**-producing strain *P. fluorescens* ATCC 39502 in the late 1980s and early 1990s allowed the proposition of a likely pathway and form the foundations for the work described in this thesis (Herbert and Knaggs, 1988; 1990; 1992a; 1992b).

Initial inspection of the structure of **1** indicated that the two aromatic rings, and at least some of the other C atoms, have their origins in the shikimic acid pathway (Knaggs, 2003). The **2** moiety is analogous to that seen in chloramphenicol (Figure 1.10),

though its side chain is distinguished in comprising one additional C atom. ^{13}C NMR analysis of initial experiments supplementing production cultures with D-[U- ^{13}C]glucose, generated coupling constants for the two aromatic rings consistent with an origin in the shikimic acid pathway via chorismate (Herbert and Knaggs, 1992a - Figure 1.20). The pattern observed for C-3 through C-10 is that expected of biosynthesis from an aromatic amino acid such as **3**, which is also a precursor in the biosynthesis of chloramphenicol and pristinamycins I and II (Blanc *et al.*, 1997; He *et al.*, 2001). Consistent with this, cultures fed L-[2,6- $^3\text{H}_2$]**3** revealed it to be an excellent precursor for **1** (10.3% specific incorporation), whereas L-[2,6- $^3\text{H}_2$]4-nitrophenylalanine and L-[2,6- $^3\text{H}_2$]4-phenylalanine were both poor precursors (Herbert and Knaggs, 1988). **1** production cultures were also found to accumulate 4-nitrophenylethanol (4-NPE; **4**) and 4-nitrophenylacetic acid (4-NPAA; **5**) as co-metabolites, suggesting they might be intermediates between **3** and **2**. However, neither [2- $^2\text{H}_2$]**4** nor [2- $^2\text{H}_2$]**5** were incorporated into **1**. The involvement of the CoA ester of **5** was also ruled out (Herbert and Knaggs, 1990). DL-[2,3- $^2\text{H}_3$]**3** was also an excellent precursor to **1**, but loss of deuterium at the C-2 position indicated that this carbon originates from an alternative source.

Glycine was identified as a prime candidate to provide the C-1 and C-2 carbons (Herbert and Knaggs, 1990). Using 'resting' *P. fluorescens* cells in MES buffer, supplementation with the precursors 2,3-DHBA (**6**), **3** and [2- ^{13}C]glycine showed the expected enhancement of the C-2 signal in ^{13}C NMR experiments and no product was observed in the absence of any of the precursors. However, enhancement of the signal for the C-1 unit was also detected and both C-1 and C-2 units had doublets associated with them. A small amount of enrichment was also detected for the C-3 and C-4 units (Herbert and Knaggs, 1990). Whilst the use of 'resting cells' was applied to minimise the effect of precursor channelling into primary metabolism, these results can be explained by the use of glycine as a growth source via the glyoxylate and tartronic semialdehyde pathways: the incorporation of two ^{13}C units into tartronic semialdehyde given the levels of [2- ^{13}C]glycine supplemented would be likely, and glyoxylate subsequently generated from acetyl-CoA in the glyoxylate pathway would be doubly labelled as a result (Figure 1.21). Similarly, labelled C-3 and C-4 signals would arise from doubly labelled phosphoenolpyruvate, the substrate of deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase which is the first enzyme in the shikimic acid pathway (Knaggs, 2003).

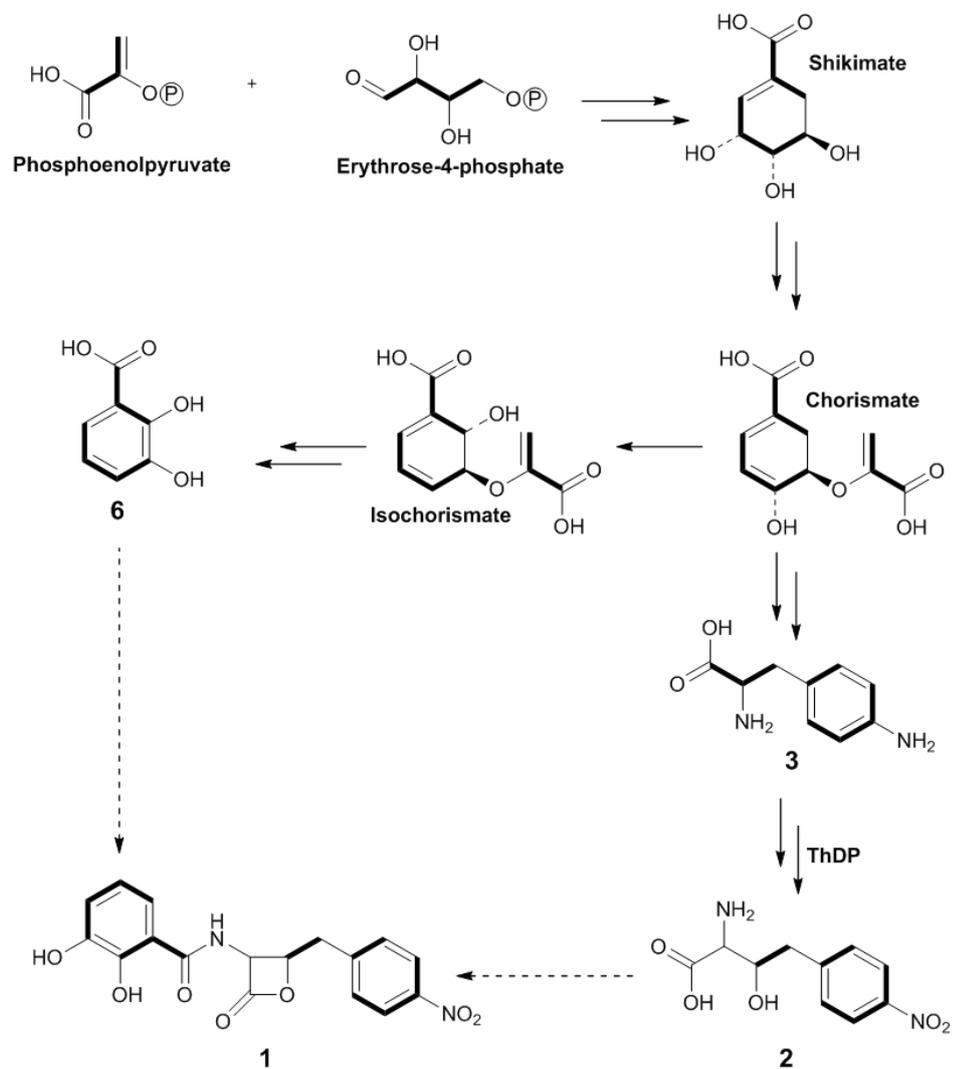


Figure 1.20. Illustration of interpreted results from *D*-[U-¹³C]glucose and *L*-[2,6-³H₂]4-aminophenylalanine feeding experiments conducted by Herbert and Knaggs (1992a). Adapted with permission from The Royal Society of Chemistry.

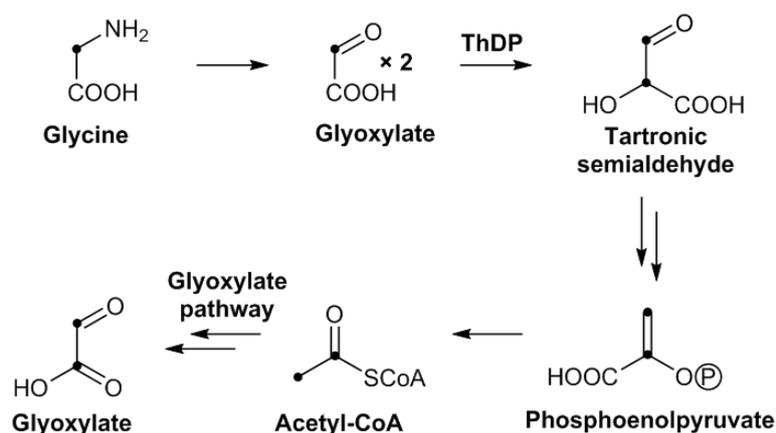


Figure 1.21. The tartronic semialdehyde pathway. ^{13}C atoms are labelled to illustrate how they can be introduced at both the C-1 and C-2 positions in glyoxylate in *P. fluorescens* cultures fed $[2-^{13}\text{C}]$ glycine, consistent with the ^{13}C NMR results observed by Herbert and Knaggs (1990; 1992).

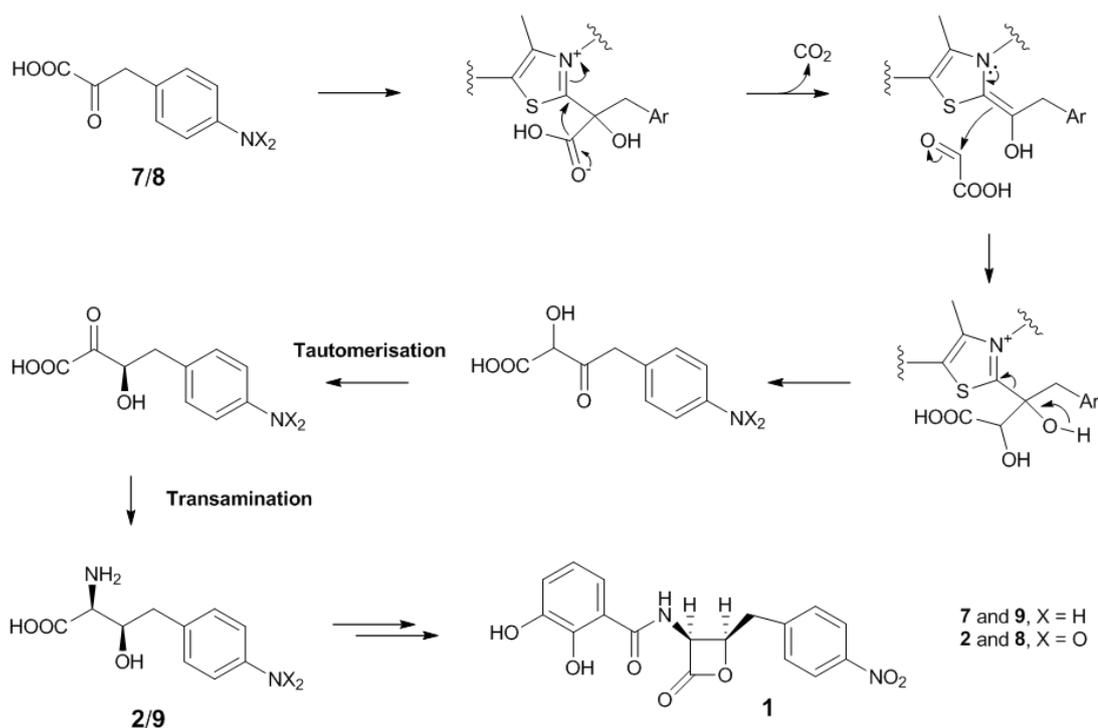


Figure 1.22. Decarboxylative ThDP-dependent mechanism for the biosynthesis of AHNB (2) proposed by Herbert and Knaggs (1990). Adapted with permission from Elsevier.

Based on their experimental data, Herbert and Knaggs proposed a mechanism for the biosynthesis of **2** involving a ThDP-mediated condensation to catalyse C-C bond formation between glyoxylate and either 4-aminophenylpyruvic acid (4-APP; **7**) or 4-nitrophenylpyruvic acid (4-NPP; **8**), the transamination product of **3** (Figure 1.22 - Herbert and Knaggs, 1990). This mechanism was favoured based on the observation of deuterium loss from the C-2 of deuterated **3**, and the existence of analogous ThDP-catalysed condensations in yeast and *Ephedra gerardiana* (Grue-Sørensen & Spenser, 1988; 1989; Shin and Rogers, 1995). The product of this anticipated ThDP-dependent condensation reaction would be an acyloin intermediate that would need to undergo further tautomerisation and transamination to generate **2** (Figure 1.22). (2*S*,3*R*)-2-amino-3-hydroxy-4-(4'-aminophenyl)butanoate (AHAB; **9**) could also be the product depending on when the nitro group of **1** is introduced. Alternative schemes proposed by Herbert and Knaggs involved 4-aminophenylacetaldehyde as the substrate instead of **7**, or the ThDP-derivative of glyoxylate reacting with either 4-aminophenylacetaldehyde or **5**, but these schemes lack supporting analogies.

1.3.4.1 Project outlook

The ThDP-dependent mechanism proposed by Herbert and Knaggs makes a mechanism-guided genome mining strategy particularly appealing because it would represent a novel pathway to β -OH- α -AAs. ThDP-dependent enzyme and transaminase queries, coupled with a query for peptide assembly (e.g. NRPS), might provide a 'fingerprint' for the identification of further NP BGCs encoding unusual nonproteinogenic amino acid precursors and novel chemistry.

Furthermore, in the absence of tautomerisation (or using a transaminase that has α -hydroxy- β -keto acyloin tautomer substrate specificity), a similar mechanism could be envisaged for the biosynthesis of α -hydroxy- β -amino acids (α -OH- β -AAs), expanding the potential chemical space accessible using a mechanism-guided approach. These building blocks are rarer in nature but when present are also associated with bioactive NPs, such as the anticancer agents bleomycin and taxol, and the cytotoxin microcystin (Kudo *et al.*, 2014). In addition to introducing multiple stereocentres, α -OH- β -AAs offer greater stability to peptides that comprise them as they are less often recognised by protease-type hydrolases. The aminopeptidase inhibitors amastatin (Aoyagi *et al.*, 1978) and bestatin (Ubenimex) (Umezawa *et al.*, 1976) represent ideal candidate NPs to validate a mechanism-guided genome mining approach, as both comprise α -OH- β -AA moieties (Figure 1.23), but neither have been characterised at the biosynthetic level. Using a ThDP-dependent mechanism-guided approach, one

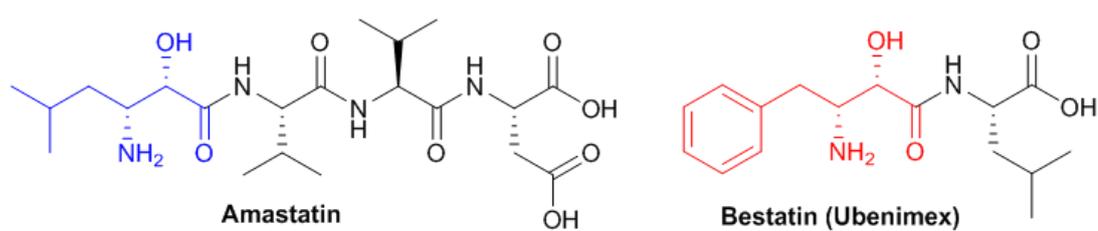


Figure 1.23. Structures of amastatin and bestatin. α -OH- β -AA moieties are highlighted in blue for amastatin and red for bestatin.

might be able to better explore the chemical space locked away in the genomes of microorganisms in a directed manner, enriching for NPs that likely comprise novel chemistry and thus may interact with novel biological targets. By discovering more of these biosynthetic enzymes, and identifying appropriate transaminase partners, a significant contribution can be made to the synthetic toolbox for the preparation of highly privileged nonproteinogenic amino acid building blocks.

1.4 Project objectives

1. Delineate the biosynthetic pathway of **1** at the molecular level
2. Biochemically characterise the biosynthetic enzyme(s) responsible for the biosynthesis of **2** and assess its potential utility as a biocatalyst
3. Biochemically characterise the mechanism of β -lactone ring formation
4. Identify the biological target of **1** and characterise its mechanism of interaction
5. Explore the potential for mechanism-guided genome mining/characterise α -OH- β -AA biosynthesis amastatin and/bestatin

Chapter 2:

Materials and methods

Chapter 2: Materials and methods

2.1 General materials

Reagents and chemicals were purchased from Alfa-Aesar, Sigma-Aldrich, Santa Cruz Biotechnology Inc., Amatek Chemical Co., Ltd., and BD Biosciences, and were used without further purification. All solvents used for HPLC were obtained from Fisher Scientific (at least of HPLC grade) and were filtered before use.

Commercially available enzymes, reagents, DNA purification kits and size markers for DNA and protein work used in this study are listed in Table 2.1.

Supplier	Enzymes and kits
Invitrogen™ (Thermo Fisher Scientific), Paisley, UK	DNA polymerase I, Large (Klenow) fragment
New England Biolabs®, Ipswich, UK	1 kb DNA ladder Colour Prestained Protein Standard, Broad Range (11-245 kDa) Gibson Assembly® Master Mix Q5® High-Fidelity DNA polymerase T4 DNA ligase
Promega, Southampton, UK	GoTaq® G2 Green Master Mix RNase ONE™ Ribonuclease Wizard® SV Gel and PCR product Clean-Up System Wizard® Plus SV Minipreps DNA purification System
Roche, Burgess Hill, UK	Deoxynucleoside Triphosphate (dNTPs) Set PCR Grade rAPid Alkaline Phosphatase

Table 2.1. Lab consumables used in this work

2.2 Bacterial strains, plasmids and oligonucleotides

2.2.1 Bacterial strains

Bacterial strains used in this study are listed in Table 2.2. All strains were maintained on solid LB medium (with appropriate selection) at 37°C (*E. coli*) or 28°C (*P. fluorescens*).

2.2.2 Plasmids

Plasmids used and constructed in this study are listed in Table 2.3.

2.2.3 Oligonucleotides and primers

Oligonucleotides and primers used in this study are listed in Supplementary Table 1 (Appendix 1).

Strain	Description/Genotype	Reference
<i>B. subtilis</i> EC 1524	Bioassay strain; <i>trpC2</i> , Subtilin BGC deleted	O'Rourke <i>et al.</i> , 2017
<i>E. coli</i> ATCC 25922	Bioassay strain; WT	ATCC, USA
<i>E. coli</i> ATCC 25922 pJH10TS- <i>obaO</i>	ATCC 29522 carrying a WT copy of the <i>obaO</i> gene as a <i>BmtI-KpnI</i> fragment cloned into pJH10TS for expression <i>in trans</i>	This work
<i>E. coli</i> DH5 α	Cloning strain; F ⁻ , <i>endA1</i> , <i>glnV44</i> , <i>thi-1</i> , <i>recA1</i> , <i>relA1</i> , <i>gyrA96</i> , <i>deoR</i> , <i>nupG</i> , ϕ 80 <i>dlac</i> Δ (<i>lacZ</i>)M15, Δ (<i>lacI</i> ZYA- <i>argF</i>)U169, <i>hsdR17</i> ($\text{rK}^- \text{mK}^+$), λ^-	Hanahan, 1983
<i>E. coli</i> NiCo21(DE3):pLysS	Expression strain; <i>can::CBD</i> , <i>fhuA2</i> , [<i>lon</i>] <i>ompT</i> , <i>gal</i> (λ DE3) [<i>dcm</i>] <i>arnA::CBD</i> , <i>slyD::CBD</i> , <i>glmS6Ala</i> , Δ <i>hsdS</i> λ DE3 = λ <i>sBamHlo</i> Δ <i>EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21</i> , Δ <i>nin5</i> . pLysS subsequently introduced	New England Biolabs
<i>E. coli</i> NiCo21(DE3):pLysS:pET28a(+)- <i>obaC</i>	NiCo21(DE3):pLysS carrying a WT copy of the <i>obaC</i> gene as an <i>NdeI-XhoI</i> fragment for production of the ObaC protein with an N-terminal hexahistidine tag in pET28a(+)	This work
<i>E. coli</i> NiCo21(DE3):pLysS:pET28a(+)- <i>obaG</i>	NiCo21(DE3):pLysS carrying a WT copy of the <i>obaG</i> gene as an <i>NdeI-XhoI</i> fragment for production of the ObaG protein with an N-terminal hexahistidine tag in pET28a(+)	This work
<i>E. coli</i> NiCo21(DE3):pLysS:pET28a(+)- <i>obaH</i>	NiCo21(DE3):pLysS carrying a WT copy of the <i>obaH</i> gene as an <i>NdeI-XhoI</i> fragment for production of the ObaH protein with an N-terminal hexahistidine tag in pET28a(+)	This work
<i>E. coli</i> NiCo21(DE3):pLysS:pET28a(+)- <i>obaI</i>	NiCo21(DE3):pLysS carrying a WT copy of the <i>obaI</i> gene as an <i>NdeI-XhoI</i> fragment for production of the Obal protein with an N-terminal hexahistidine tag in pET28a(+)	This work
<i>E. coli</i> NiCo21(DE3):pLysS:pET28a(+)- <i>obaI</i> C1141A	NiCo21(DE3):pLysS carrying a copy of the <i>obaI</i> gene with a C1141A substitution as an <i>NdeI-XhoI</i> fragment for production of the Obal C1141A protein with an N-terminal hexahistidine tag in pET28a(+)	This work
<i>E. coli</i> NiCo21(DE3):pLysS:pET28a(+)- <i>obaK</i>	NiCo21(DE3):pLysS carrying a WT copy of the <i>obaK</i> gene as an <i>NdeI-XhoI</i> fragment for production of the ObaK protein with an N-terminal hexahistidine tag in pET28a(+)	This work
<i>E. coli</i> NiCo21(DE3):pLysS:pET28a(+)- <i>thrRS</i>	NiCo21(DE3):pLysS carrying a WT copy of the <i>E. coli thrRS</i> gene as an <i>NdeI-XhoI</i> fragment for production of the ThrRS protein with an N-terminal hexahistidine tag in pET28a(+)	This work
<i>E. coli</i> NiCo21(DE3):pLysS:pET28a(+)- <i>thrRS</i> C480V	NiCo21(DE3):pLysS carrying a copy of the <i>E. coli thrRS</i> gene with a C480V substitution as an <i>NdeI-XhoI</i> fragment for production of the ThrRS C480V protein with an N-terminal hexahistidine tag in pET28a(+)	This work
<i>E. coli</i> NR698	Bioassay strain; MC4100 (F ⁻ <i>araD139</i> Δ (<i>argF-lac</i>)U169, <i>rpsL150</i> , <i>relA1</i> , <i>flbB5301</i> , <i>deoC1</i> , <i>ptsF25</i> , <i>rbsR</i>), <i>imp4213</i>	Ruiz <i>et al.</i> , 2005
<i>E. coli</i> NR698:pJH10TS- <i>obaO</i>	NR698 carrying a WT copy of the <i>obaO</i> gene as a <i>BmtI-KpnI</i> fragment cloned into pJH10TS for expression <i>in trans</i>	This work

<i>E. coli</i> S17-1 λ (pir)	Conjugation strain; <i>recA</i> , <i>thi</i> , <i>pro</i> , <i>hsd</i> (R ⁻ M ⁺)RP4: 2- Tc::Mu- Km::Tn7 λ pir SM ^R Tp ^R	Simon <i>et al.</i> , 1983
<i>P. fluorescens</i> ATCC 39502	Obafluorin-producing strain, WT	ATCC, USA
<i>P. fluorescens</i> Δ <i>obaC</i>	ATCC 39502 with an in-frame truncation in the <i>obaC</i> gene	This work
<i>P. fluorescens</i> Δ <i>obaC</i> :pJH10TS- <i>obaC</i>	Δ <i>obaC</i> strain carrying a WT copy of the <i>obaC</i> gene as a <i>BmtI-KpnI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work
<i>P. fluorescens</i> Δ <i>obaF</i>	ATCC 39502 with an in-frame truncation in the <i>obaF</i> gene	This work
<i>P. fluorescens</i> Δ <i>obaF</i> :pJH10TS- <i>obaF</i>	Δ <i>obaF</i> strain carrying a WT copy of the <i>obaF</i> gene as a <i>BmtI-KpnI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work
<i>P. fluorescens</i> Δ <i>obaG</i>	ATCC 39502 with an in-frame truncation in the <i>obaG</i> gene	This work
<i>P. fluorescens</i> Δ <i>obaG</i> :pJH10TS- <i>obaG</i>	Δ <i>obaG</i> strain carrying a WT copy of the <i>obaG</i> gene as a <i>BmtI-KpnI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work
<i>P. fluorescens</i> Δ <i>obaH</i>	ATCC 39502 with an in-frame truncation in the <i>obaH</i> gene	This work
<i>P. fluorescens</i> Δ <i>obaH</i> :pJH10TS- <i>obaH</i>	Δ <i>obaH</i> strain carrying a WT copy of the <i>obaH</i> gene as a <i>BmtI-XbaI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work
<i>P. fluorescens</i> Δ <i>obaI</i>	ATCC 39502 with an in-frame truncation in the <i>obaI</i> gene	This work
<i>P. fluorescens</i> Δ <i>obaI</i> :pJH10TS- <i>obaI</i>	Δ <i>obaI</i> strain carrying a WT copy of the <i>obaI</i> gene as a <i>BmtI-XbaI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work
<i>P. fluorescens</i> Δ <i>obaJ</i>	ATCC 39502 with an in-frame truncation in the <i>obaJ</i> gene	This work
<i>P. fluorescens</i> Δ <i>obaJ</i> :pJH10TS- <i>obaJ</i>	Δ <i>obaJ</i> strain carrying a WT copy of the <i>obaJ</i> gene as a <i>BmtI-KpnI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work
<i>P. fluorescens</i> Δ <i>obaK</i>	ATCC 39502 with an in-frame truncation in the <i>obaK</i> gene	This work
<i>P. fluorescens</i> Δ <i>obaK</i> :pJH10TS- <i>obaK</i>	Δ <i>obaK</i> strain carrying a WT copy of the <i>obaK</i> gene as a <i>BmtI-KpnI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work
<i>P. fluorescens</i> Δ <i>obaL</i>	ATCC 39502 with an in-frame truncation in the <i>obaL</i> gene	This work
<i>P. fluorescens</i> Δ <i>obaL</i> :pJH10TS- <i>obaL</i>	Δ <i>obaL</i> strain carrying a WT copy of the <i>obaL</i> gene as a <i>BmtI-XbaI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work
<i>P. fluorescens</i> Δ <i>obaM</i>	ATCC 39502 with an in-frame truncation in the <i>obaM</i> gene	This work

<i>P. fluorescens</i> Δ obaM:pJH10TS-obaM	Δ obaM strain carrying a WT copy of the <i>obaM</i> gene as a <i>BmtI-XbaI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work
<i>P. fluorescens</i> Δ obaO	ATCC 39502 with an in-frame truncation in the <i>obaO</i> gene	This work
<i>P. fluorescens</i> Δ obaO:pJH10TS-obaO	Δ obaO strain carrying a WT copy of the <i>obaO</i> gene as a <i>BmtI-KpnI</i> fragment cloned into pJH10TS for complementation by expression <i>in trans</i>	This work

Table 2.2. Strains used in this work

Plasmid	Genotype/description	Reference
pET28a(+)	Expression vector; Kan ^R . the transcription of the cloned gene is driven by the T7 RNA polymerase and controlled by the LacI repressor, <i>ColE1</i> replicon	Novagen
pME3087	Suicide vector; <i>ColE1</i> replicon, <i>IncP-1</i> , <i>Mob</i> ; Tc ^R	Voisard <i>et al.</i> , 1994
pTS1	pME3087 modified with a <i>sacB</i> gene for counter-selection and an expanded multiple cloning site	This work
pJH10	Vector for complementation studies by <i>in trans</i> expression in <i>P. fluorescens</i> ATCC 39502 Δ strains; pOLE1 with <i>IncC1</i> deleted, <i>EcoRI-SacI</i> polycloning site, Tc ^R from pDM1.2	El-Sayed <i>et al.</i> , 2003
pJH10TS	pJH10 modified with an expanded cloning site	This work
pLysS	Vector for basal expression from the T7 promoter by producing T7 lysozyme; p15A replicon, Cm ^R	Novagen

Table 2.3. Plasmids used in this work

2.3 Culture media, buffers and solutions

2.3.1 Culture media

Where solid equivalents of media were used, agar 10 g/L was additionally added to the recipes presented below.

Luria Bertani (LB)

Difco™ Bacto tryptone	10 g/L
Difco™ yeast extract	10 g/L
NaCl	5 g/L
Glucose	1 g/L

Luria Bertani (LB) 7% sucrose

Difco™ Bacto tryptone	10 g/L
Difco™ yeast extract	10 g/L
NaCl	5 g/L
Glucose	1 g/L
Sucrose	7 g/L

Müller Hinton (MH) broth

Beef infusion solids	2 g/L
Casein hydrolysate	17.5 g/L
Starch	1.5 g/L

Terrific broth (TB)

Difco™ Bacto tryptone	12 g/L
Difco™ yeast extract	24 g/L
Glycerol	4 mL/L

1 production medium (OPM)

Difco™ yeast extract	5 g/L
D-Glucose	5 g/L
FeSO ₄	0.1 g/L
MgSO ₄	0.1 g/L

SOC medium

Tryptone	20 g/L
Yeast extract	5 g/L
NaCl	0.58 g/L
KCl	0.19 g/L
MgCl ₂	2 g/L
MgSO ₄	2.5 g/L

Soft Nutrient Agar (SNA)

Difco™ Nutrient Broth	8 g/L
Agar	7 g/L

2.3.2 Buffers and solutions**4-(2-pyridylazo)-resorcinol (PAR) solution**

PAR	7.1 mg/L	30 μM
PBS (pH 7.0)		

Chrome azurol S (CAS) assay solution

Prepared as described by Alexander and Zuberer (1990). 21.9 mg HDTMA (601 μ M final conc.) was dissolved in 25 mL ddH₂O while stirring constantly over low heat. 1.5 mL of 1 mM FeCl₃·6H₂O (15 μ M final conc., dissolved in 10 mM HCl) was mixed with 7.5 mL of 2 mM CAS (150 μ M final conc.). The resulting solution was gradually added to the HDTMA with stirring. 9.76 g MES (500 mM final conc.) was dissolved in 50 mL ddH₂O pH 5.6 (adjusted with 50% KOH) and applied to the dye solution. ddH₂O was added to a final volume of 100 mL.

***E. coli* threonyl-tRNA synthetase (ThrRS) purification buffer**

HEPES/HCl (pH 7.8)	6 g/L	25 mM
NaCl	17.5 g/L	300 mM
MgCl ₂	952 mg/L	10 mM
Glycerol	100 mL/L	10 % (v/v)
Imidazole*	17 g/L	250 mM

Loading dye for DNA samples (6× concentrated)

Bromophenol blue	2.5 g/L	3.7 mM
Xylene cyanol FF	2.5 g/L	4.6 mM
Glycerol	165 mL/L	33% (v/v)

Loading dye for protein samples (4× concentrated)

Tris/HCl (pH 8.0)	29 g/L	240 mM
SDS (25%)	80 g/L	8% (v/v)
Bromophenol blue	0.4 g/L	0.04% (w/v)
β -mercaptoethanol	40 mL/L	4% (v/v)
EDTA	14.6 g/L	50 mM
Glycerol	400 mL/L	40% (v/v)

ObaG buffer

HEPES/HCl (pH 7.8)	6 g/L	25 mM
NaCl	17.5 g/L	300 mM
MgCl ₂	952 mg/L	10 mM
PLP	100 mg/L	0.4 mM
Glycerol	100 mL/L	10 % (v/v)
Imidazole*	17 g/L	250 mM

ObaH buffer

HEPES/HCl (pH 7.8)	6 g/L	25 mM
NaCl	17.5 g/L	300 mM
MgCl ₂	952 mg/L	10 mM
ThDP	230 mg/L	0.5 mM
Glycerol	100 mL/L	10 % (v/v)
Imidazole*	17 g/L	250 mM

ObaCIK purification buffer

Tris/HCL (pH 7.8)	6 g/L	
NaCl	17.5 g/L	300 mM
MgCl ₂	952 mg/L	10 mM
Glycerol	100 mL/L	10 % (v/v)
Imidazole*	17 g/L	250 mM

PLP-dependence assay buffer

HEPES/HCl (pH 7.8)	6 g/L	25 mM
NaCl	17.5 g/L	300 mM
MgCl ₂	952 mg/L	10 mM

Running buffer for SDS-PAGE

Tris/HCl (pH 8.0)	3 g/L	25 mM
Glycine	14.4 g/L	192 mM
SDS (25%)	1 g/L	0.1% (w/v)

Tris-Borate-EDTA (TBE) buffer, pH 8.3

Trizma base	10.8 g/L	89 mM
Boric acid	5.5 g/L	89 mM
EDTA	930 mg/L	3 mM

Tris-EDTA (TE) buffer, pH 8

Tris/HCl (pH 8.0)	1.2 g/L	10 mM
EDTA	292 mg/L	1 mM

*Imidazole only included in buffer for elution during Ni affinity chromatography (Section 2.6.1). Purified proteins were subsequently exchanged into their respective buffers without imidazole for storage at -80°C.

2.3.3 Antibiotics

The stock and working concentrations of antibiotics used in this study are listed in Table 2.4.

Antibiotic	Abbr.	Solvent	Stock concentration (mg/mL)	Working concentration (µg/mL)
Carbenicillin	Cb	ddH ₂ O	100	100
Chloramphenicol	Cm	Ethanol	25	25
Kanamycin	Kan	ddH ₂ O	50	50
Nitrofurantoin	N	DMSO	50	100
Tetracycline	Tc	Ethanol	50	25
Streptomycin	Sm	H ₂ O	100	100

Table 2.4. Stock and working concentrations of antibiotics used in this work.

2.4 Cloning, screening and transformations

2.4.1 Preparation of *P. fluorescens* ATCC 39502 genomic DNA

10 mL LB cultures were inoculated from a -80°C 40% glycerol stock and were incubated O/N at 28°C, 250 rpm. 2 mL of culture were pelleted at 13,200 rpm for 2 min in a 2 mL eppendorf and the supernatant was subsequently discarded. Cells were re-suspended in 600 µL 3 M guanidinium thiocyanate/10 mM EDTA (pH 7.0) and incubated at 85°C for 5 min, before being allowed to cool back down to room temperature. 1 µL RNase ONE™ and 60 µL of the corresponding 10× buffer were then added and inverted 25 times to mix, before incubation at 37°C for 15 min. 200 µL of 10 M ammonium acetate (pH 7.0) solution was then applied and mixed by thorough inversion, before centrifugation at 13,200 rpm for 3 min. 600 µL of supernatant were then removed and 600 µL isopropanol were introduced. Eppendorfs were then inverted 50 times to precipitate the DNA, before centrifugation at 13,200 rpm for 1 min and removal of supernatant. 400 µL of 70% ethanol were then added to wash the DNA by inversion. Supernatant was removed as before and the ethanol wash was repeated once more. Supernatant was removed following centrifugation and samples were left to air dry to remove residual ethanol. Genomic DNA was then re-suspended in TE buffer by incubation at 65°C with intermittent flicking for 5 h. Genomic DNA samples were aliquoted and stored either at -20 or 4°C.

2.4.2 Polymerase chain reaction (PCR)

All PCR reactions were performed in a T100™ Thermal Cycler (Bio-Rad). PCR products for cloning and/or sequencing were generated using the high-fidelity Q5 polymerase in 25 µL reaction volumes, following the manufacturer's instructions (Table 2.5). *P. fluorescens* ATCC 39502 genomic DNA was used as template. Colony

PCR to screen colonies for introduced gene deletions or to verify constructs was performed using GoTaq polymerase in 10 μ L reaction volumes (Table 2.6). A toothpick was used to apply cells to 50 μ L dH₂O which was used as template. PCR products were purified using a Wizard® SV Gel and PCR product Clean-Up System (Promega).

High-fidelity PCR reaction mix		
Component	Volume (μ L)	Final concentration
5x Q5 reaction buffer	5	1x
Forward primer	1.25	0.5 μ M
Reverse primer	1.25	0.5 μ M
DMSO	0.75	0.03%
dNTPs	0.5	200 μ M
DNA template	0.5	~50 ng
Q5 polymerase	0.25	0.5 U

Q5 PCR programme			
Cycle	Temperature ($^{\circ}$ C)	Time	Repeat (x)
Initial denaturation	98	3 min	1
Denaturation	98	30 s	30
Annealing	55-65	30 s	
Elongation	72	30 s/kb	
Final elongation	72	10 min	1
Cooling	12	1 min	1

Table 2.5. High-fidelity PCR reaction mix and programme

Colony PCR reaction mix		
Component	Volume (μ L)	Final concentration
GoTaq G2 Master Mix	5	1x
Forward primer	0.5	1.0 μ M
Reverse primer	0.5	1.0 μ M
DMSO	0.5	1%
DNA template	3.5	Cell solution

Colony PCR programme			
Cycle	Temperature ($^{\circ}$ C)	Time	Repeat (x)
Initial denaturation	95	5 min	1
Denaturation	95	30 s	30
Annealing	55-65	30 s	
Elongation	72	1 min/kb	
Final elongation	72	10 min	1
Cooling	12	1 min	1

Table 2.6. Colony PCR reaction mix and programme

2.4.3 Agarose gel electrophoresis

Agarose gels were prepared by adding agarose to TBE buffer to a final concentration between 0.8 and 1.2%, depending on the size of DNA fragments for analysis. 10 μ g/mL ethidium bromide was added before pouring for subsequent visualisation under ultraviolet (UV) light. DNA was prepared by addition of loading dye before gel

application. 1 kb ladder (NEB) was applied in parallel for analysis. Gel electrophoresis was performed in TBE buffer using a PowerPac™ Universal Power Supply (Bio-Rad) run between 80-120 V, until DNA fragments were deemed to be sufficiently separated. Gel visualisation was performed using a UV Transilluminator and Gel Documentation System (UVP).

2.4.4 Cloning of DNA fragments

Cloning was performed following standard molecular biology procedures. Vector DNA or PCR amplified DNA fragments were cut with the appropriate restriction endonucleases according to the manufacturer's instructions. Restriction digest mixes were performed in 50 µL reaction volumes (20 µL for analytical digests) and were incubated at 37°C for 2 h. Vector digests were subsequently treated with alkaline phosphatase (addition of 1 µL alkaline phosphatase, 6 µL alkaline phosphatase buffer and 3 µL ddH₂O) and incubated for a further 45 min at 37°C. Digests were either purified directly, or from size-separated gel fragments following electrophoresis, using a Wizard® SV Gel and PCR product Clean-Up System. Ligations were performed O/N at 8°C using 17 µL digested DNA, 1 µL T4 ligase and 2 µL 10× T4 ligase buffer. Ligation mixes were directly used for transformation of chemically competent *E. coli* DH5α cells.

2.4.5 Restriction site removal

To remove an unwanted *Nde*I site from the pTS1 suicide vector, a digest of the vector with *Nde*I was performed as above but incubation was at 37°C for 30 min only. Singly cut plasmid was gel-purified before the subsequent reaction was set up and incubated at room temperature for 20 min to 'fill in' the unwanted restriction site:

4 µL	React buffer 10× (Invitrogen™)
1.5 µL	dNTPs (10 mM final conc.)
1.5 µL	DNA polymerase I (Klenow) (Invitrogen™) (10 U final conc.)
20 µL	Partially digested pTS1
13 µL	ddH ₂ O

A standard ligation as described previously was then performed to re-circularise blunt-ended vector and ligation mixes were directly used for transformation of chemically competent *E. coli* DH5α cells. An analytical digest was performed to identify clones in which the correct *Nde*I site had been removed.

2.4.6 Gibson Assembly

In order to create an Obal expression construct with a C1141A substitution (TGC→GCC) Gibson Assembly® (Gibson *et al.*, 2009) was used as it allowed the introduction of base substitutions in overlapping sequences. The Obal gene was cloned as two overlapping fragments, that overlapped in the region where C1141 was encoded (Appendix 1 - Supplementary Table 1). This allowed identical double bp substitutions to be introduced in the primer sequences for each fragment in this region, which were introduced following the ligation of both fragments together along with the pET28a(+) backbone (also amplified as a PCR fragment with which both fragments' primers overlap). A similar approach was applied for C480V substitutions in *E. coli* threonyl-tRNA (ThrRS) synthetase constructs. Primers were designed using NEBuilder® Assembly Tool (NEB) and PCR was performed with Q5 polymerase as described previously. PCR products were purified using a Wizard® SV Gel and PCR product Clean-Up System (Promega) and product concentrations were calculated using a NanoDrop-1000 (Thermo Fisher Scientific). Reactions were assembled following manufacturer's instructions in containing 10 µL of 2×Gibson Assembly Master Mix (NEB) and a suitable ratio of PCR products, made up with ddH₂O to a final reaction volume of 20 µL. Samples were then incubated at 50°C for 15 min, before transformation of the ligation mix into chemically-competent *E. coli*.

2.4.7 Preparation and transformation of chemically competent *E. coli*

A 5 mL LB culture was established for the *E. coli* strain to be transformed, grown O/N at 37°C, 250 rpm. 500 µL was used to inoculate a 100 mL LB culture which was grown at 37°C, 250 rpm until OD_{600 nm} = 0.3-0.4. Cells were centrifuged at 4,500 rpm for 10 min at 4°C, before gentle resuspension in 12.5 mL of ice-cold 100 mM MgCl₂. The resulting cell suspension was centrifuged for 10 min at 4,000 rpm, 4°C. Cells were then re-suspended in 25 mL ice-cold 100 mM CaCl₂ and incubated on ice for ~1 h. Cell suspensions were then pelleted at 4,000 rpm for 10 min at 4°C, before re-suspension in 1 mL ice-cold 100 mM CaCl₂ in 20% glycerol (v/v). 200 µL aliquots were made and snap-frozen with liquid nitrogen for storage at -80°C.

To transform chemically competent cells, they were first thawed on ice. 50 µL cells were mixed with either 5 µL of ligation mix, 2 µL Gibson Assembly mix or 0.5 µL of vector DNA and incubated on ice for 30 min. Cells were then heat shocked at 42°C for 25 s before returning to ice for a further 3 min. 950 µL of SOC medium were then added to the transformed cells and recovery was performed at 37°C, 250 rpm for ~1 h. Cells were finally plated on LB agar with appropriate selection and incubated O/N

at 37°C. Resulting colonies were screened by colony PCR to identify successful colonies.

2.4.8 Plasmid purification and sequencing

5 mL *E. coli* LB cultures with appropriate selection were established from single colonies on plates following cloning or from -80°C 40% glycerol stocks, and were grown O/N at 37°C, 250 rpm. Plasmid DNA was purified using a Wizard® Plus SV Minipreps DNA purification System (Promega). All constructs (and PCR products where necessary) generated during this project were sequenced by Sanger sequencing (Eurofins Genomics).

2.5 Mutagenesis and complementation experiments

2.5.1 pTS1 and pJH10TS construct design

For pTS1-mediated knockout constructs, primers were designed to amplify 800-1,200 bp flanking regions of a selected *oba* PCS for deletion, for cloning as *Xba*I-*Avr*II and *Avr*II-*Bmt*I fragments into pTS1. Flanking regions were designed to contain 10-50 PCS codons at either end of the *oba* locus for disruption to minimise possible polar effects. This would create a truncated copy of the *oba* PCS with an in-frame deletion and internal *Avr*II site, an artefact from ligation of the two flanking fragments. This would then be used to replace the WT chromosomal copy of the gene by double homologous recombination. To complement mutations, WT copies of *oba* PCSs were cloned from start to stop codon as *Bmt*I-*Kpn*I/*Xba*I fragments into pJH10TS, for introduction and ectopic expression in the relevant mutant background.

2.5.2 Conjugation of constructs into *P. fluorescens* ATCC 39502

Constructs were first transformed into *E. coli* S17-1, using empty vector constructs as negative controls. 5 mL O/N LB cultures with appropriate selection were prepared for the *E. coli* S17-1 donor and *P. fluorescens*. Cultures were centrifuged at 13,200 rpm, supernatant was discarded, and the pellets were washed twice in sterile LB medium, pre-warmed to 50°C. Pellets were ultimately re-suspended in 1.5 mL LB and were heat shocked for 15 min (*E. coli* - 37°C, *P. fluorescens* - 45°C). Conjugation suspensions were then prepared by mixing 700 µL of *P. fluorescens* cells with 300 µL *E. coli* S17-1 donor cells. Suspensions were re-suspended in 50 µL sterile LB, following centrifugation at 13,200 rpm for 1 min, and pipetted to the centre of a 0.22 µM GSTF mixed cellulose ester filter (Merck) placed on an LB agar plate. Plates were incubated at 28°C for 5 h. Cells were then re-suspended in 2 mL phosphate buffered saline (PBS) and were plated onto LB Tc²⁵ N¹⁰⁰ (to select for the construct and against

E. coli respectively), following serial dilution. Colonies were then subsequently screened by PCR, and ultimately by Sanger sequencing.

2.5.3 pTS1 mutagenesis

Following conjugation, selection on LB Tc²⁵ N¹⁰⁰ identified single crossover mutants in which the entire pTS1 construct vector backbone is integrated into the *P. fluorescens* chromosome because pTS1 is not a self-replicating vector. Colony PCR was used to identify candidate single crossover mutants and these were used to inoculate O/N LB cultures without antibiotics, grown at 28°C, 250 rpm. O/N cultures were serially diluted and plated on LB 7% sucrose N¹⁰⁰ agar plates to select for a second crossover event. Resulting colonies were picked and patched onto LB 7% sucrose N¹⁰⁰ plates with and without Tc²⁵ to identify sensitive colonies that represent WT revertant or genuine Δoba strains. Colony PCR was then performed to verify Δoba strains and these were confirmed by Sanger sequencing of deletion PCR products. 40% glycerol stocks for three separate verified clones per gene deletion were prepared and stored at -80°C.

2.5.4 Genetic and chemical complementation of Δoba strains

pJH10TS complementation constructs were introduced into the relevant Δ backgrounds using the conjugation protocol described previously (2.5.2). Chemical complementation was achieved where possible by exogenous introduction of compounds when production cultures were established. Compounds were introduced prior to culture initiation and were dissolved in either dH₂O or DMSO where appropriate. Concentrations used for each compound fed are specified within the relevant Chapters.

2.5.5 Sample preparation for analysis of metabolite production

WT and recombinant *P. fluorescens* ATCC 39502 strains were grown in OPM. A toothpick was used to inoculate 100 ml of OPM seed culture (250 ml Erlenmeyer flask) from a 40% glycerol stock (stored at -80°C), with subsequent growth for 24 h at 25°C, 300 rpm. 1 ml of this culture was used to inoculate a 100 ml (500 ml flask) OPM production culture which was incubated under the same conditions for 14 h. Samples were prepared for HPLC/LCMS analyses by extracting 1 ml of culture broth with an equal volume of ethyl acetate with mixing at 1,400 rpm for 15 min. Samples were then centrifuged (13,200 rpm for 15 min), and the organic phase was collected and evaporated. The resulting extract was dissolved in 500 μ L MeCN and centrifuged at 13,200 rpm for 20 min to remove any remaining cell debris.

2.6 Expression and purification of recombinant proteins

2.6.1 Protein expression and purification

E. coli NiCo21(DE3):pLysS strains carrying pET28a(+)-*obaC*, pET28a(+)-*obaG*, pET28a(+)-*obaH*, pET28a(+)-*obaI* and pET28a(+)-*obaK* were cultivated in TB medium at 28°C and 250 rpm on a rotary shaker until $A_{600\text{ nm}} \sim 0.5$. Protein expression was induced by addition of 0.1 mM IPTG and incubation continued at 18°C and 200 rpm for 18 h. Cells were pelleted at 6,000 rpm and 4°C, and were subsequently re-suspended in their respective purification buffers. After disruption with an EmulsiFlex-B15 high pressure homogeniser (Avestin, Inc.), cells were pelleted at 15,000 rpm and 4°C for 30 min. The lysed supernatant was incubated with chitin resin (NEB) with gentle mixing for 30 min to remove any endogenous *E. coli* metal-binding proteins. Eluted samples were loaded onto a HisTrap excel (GE Healthcare) Ni-NTA column using an ÄKTA pure (GE Healthcare) system. Proteins were washed in 5 CV of their respective buffers containing 10, 20, 30 and 50 mM imidazole concentrations (His₆-Obal washed up to 30 mM imidazole only). His₆-Obal was eluted with 20 CV of 50 mM imidazole, and His₆-ObaC/G/H and K were eluted in 20 CV of 250 mM imidazole, collected in 2 ml fractions. Oba protein-containing fractions were combined and applied to Amicon columns (30 kDa MWCO) and diluted >1,000× to remove imidazole, before being concentrated. The His₆-Obal sample was further purified by size exclusion over a HiLoad 16/600 Superdex 200 pg column (GE Healthcare). Eluted His₆-Obal-containing fractions were combined and concentrated in an Amicon column (30 kDa MWCO). Protein samples were stored at 4°C for *in vitro* assays and long-term storage was at -80°C in their respective buffers supplemented with 20% glycerol.

2.6.2 Protein concentration quantification

Coomassie Brilliant Blue G-250 will shift from red colour under acidic conditions to blue when bound to cationic or non-polar, hydrophobic amino acid residues. This shift in absorbance can be used to quantify protein concentration using a simple colorimetric assay (Bradford, 1976). Bovine serum albumin (BSA) was used for calibration at a range of final concentrations between 1-20 µg/mL. 10 µL aliquots of purified protein sample and dilutions of BSA, were mixed with 200 µL of Protein Assay Dye Reagent (Bio-Rad) and were incubated for 5 min in the dark. The absorption at $A_{595\text{ nm}}$ was determined by comparison with a protein-free control. Protein concentrations were calculated using the linear equation generated by the BSA dilution series.

2.6.3 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were analysed by size-separation under denaturing conditions with SDS-PAGE. Protein samples were prepared by adding loading buffer to protein samples in a 1:3 ratio, up to 20 μL final volume, and heat-denaturing them at 100°C for 3 min. Samples were then loaded onto a 12% RunBlue (Expedeon Ltd.) polyacrylamide gel, along with 6 μL of Color Prestained Protein Standard (NEB) for comparison. Electrophoresis was performed at 100-180 V in SDS-PAGE running buffer using an XCell SureLock™ Mini-Cell electrophoresis system (Thermo Fisher Scientific). Gels were subsequently stained using InstantBlue (Expedeon Ltd.) for 1 h, followed by de-staining in H_2O for 1 h.

2.6.4 Confirmation of protein identity by gel purification, trypsin digest and MALDI-TOF/TOF analysis

Following SDS-PAGE, protein bands were cut out and prepared for trypsin digestion. Gel fragments were first de-stained by washing in 30% ethanol at 65°C for 15 min. Washes were repeated until gel fragments appeared clear. Ethanol was then removed and fragments were subsequently washed with 50% MeCN/50 mM TEAB for 15 min. Fragments were then incubated in 10 mM DTT/50 mM TEAB solution for 30 min at 55°C. DTT/TEAB solution was then removed and gel fragments were incubated with 30 mM iodoacetamide/50 mM TEAB solution in the dark for 30 min following a brief vortex. Iodoacetamide/TEAB was subsequently removed and the fragments were washed with 50% MeCN/50 mM TEAB for 15 min. One final wash step was performed with 50 mM TEAB before gel fragments were sliced into cubes roughly 1 mm^3 in size. Gel cubes were placed into a low-bind Eppendorf tube and washed as before with 50% MeCN/50 mM TEAB. Gel cubes were then washed with 100% MeCN, causing gel fragments to shrink and become white. Residual MeCN was removed and gel slices were supplied to Dr. Gerhard Saalbach for trypsin digestion according to standard procedures adapted from Shevchenko *et al.* (1996). The tryptic digest peptide fragments were analysed by mass spectrometry to further confirm protein identity, using an autoflex Speed MALDI-TOF/TOF mass spectrometer (Bruker Daltonics GmbH).

2.7 In vitro Oba enzyme assays

2.7.1 His₆-ObaH decarboxylase activity assay

Reactions were performed in 100 μL reaction volumes comprising 50 mM His₆-ObaH and 10 mM 4-NPP (**8**) or phenylpyruvate, all in ObaH buffer. Reactions were initiated

by introduction of the enzyme and were incubated at 27°C with 700 rpm for 5 min. A boiled enzyme sample was used as a negative control. Reactions were terminated by addition of ice-cold MeOH (100 µL) and were incubated at -20°C for 1 h to ensure full enzyme precipitation. Precipitated enzyme was pelleted at 13,200 rpm for 30 min before analysis by HR-LCMS and HPLC.

2.7.2 His₆-ObaG transaldolase activity assays, time course and kinetic analysis

Reactions were performed in 100 µL reaction volumes comprising 50 mM His₆-ObaG, 10 mM glycine/L-serine/L-threonine or [U-¹³C₄,¹⁵N]L-threonine (98% isotopic purity) (for subsequent NMR experiments), and 10 mM 4-NPA (**11**), all in ObaG buffer. Reactions were initiated by introduction of the enzyme and were incubated at 27°C and 700 rpm for 2 h. A boiled enzyme sample was used as a negative control. Reactions were terminated as described previously. HPLC and HR-LCMS analysis was performed as for the His₆-ObaH decarboxylase activity assay. Time course data were similarly collected by terminating the reaction at a range of time points up to 5 h and all time points were assayed in triplicate.

The amenability of His₆-ObaG to alternative substrates was explored by performing the assay described above but using 10 mM of either benzaldehyde or phenylacetaldehyde instead of **11** as a co-substrate with L-threonine. The reverse reaction to generate **11** and L-threonine using AHNB (**2**) with and without acetaldehyde as substrates with His₆-ObaG was also performed as above but using 20 mM concentrations of starting substrates. Single-substrate kinetic analysis was carried out by performing the transaldolase activity assay with varying concentrations of L-threonine (1–200 mM). Reactions were performed using 25 mM enzyme and were incubated at 27°C, 700 rpm for 4 min, before quenching, sample processing and HPLC analysis. Five replicates were performed for each concentration of L-threonine assayed, and data were fitted to the Michaelis-Menten equation using GraphPad Prism 5.04 (GraphPad Software, Inc., La Jolla, USA).

2.7.3 Coupled His₆-ObaH and His₆-ObaG activity assay

His₆-ObaG was first exchanged into His₆-ObaH buffer using an Amicon column (30 kDa MWCO) to avoid reaction of **8** with unbound PLP. Reactions were performed in 100 µL reaction volumes comprising 50 mM His₆-ObaG, 50 mM His₆-ObaH, 10 mM **8** and 10 mM L-threonine, all in His₆-ObaH buffer. Reactions were initiated by introduction of the enzyme and were incubated at 27°C and 700 rpm for 2 h. Control reactions were also performed in which one or both of His₆-ObaG and His₆-ObaH

were boiled before the reaction. Reactions were terminated as described previously. HPLC and HR-LCMS analysis was performed as for the His₆-ObaH decarboxylase activity assay.

2.7.4 His₆-ObaG L-penicillamine competition assay

Recombinant His₆-ObaG was incubated with L-penicillamine as described previously for serine palmitoyltransferase by Lowther *et al.* (2012). Excess PLP was removed from His₆-ObaG samples by exchanging into PLP-dependence assay buffer using an Amicon column (30 kDa MWCO). Reactions were performed in 600 µL reaction volumes comprising 20 mM His₆-ObaG and were initiated by addition of L-penicillamine (dissolved in reaction buffer) to a final concentration of 10 mM. UV spectra were recorded over 30 min using a Lambda 35 UV/Visible Spectrophotometer (PerkinElmer).

2.7.5 His₆-ObaG sodium borohydride reduction assay

His₆-ObaG samples in which excess PLP was removed, were treated with NaBH₄ to reduce the His₆-ObaG-PLP aldimine to form an amine adduct (Chen and Frey, 2001). Reactions were performed on ice for 15 min in 600 µL volumes comprising 20 mM His₆-ObaG and were initiated by addition of NaBH₄ (dissolved in reaction buffer) to a final concentration of 1 mM. Samples were analysed by UV/Visible spectrophotometry as described before.

2.7.6 His₆-Obal hydroxylamine-trapping assay

Reactions were performed as described by Kadi and Challis (2009). Reaction mixtures comprised 8.5 mM His₆-Obal, 50 mM Tris/HCl pH 8.0, ATP (2.25 mM), 150 mM hydroxylamine, 5 mM acid substrate and 15 mM MgCl₂ in a final volume of 300 µL, and were allowed to proceed at 28°C for 5 h. Boiled enzyme and no substrate reactions were performed as negative controls. Following reaction termination by addition of quenching solution (10% FeCl₃ 6H₂O and 3.3% trichloroacetic acid made up in 0.7 M HCl), and centrifugation to pellet precipitated protein at 13,200 rpm for 15 min, samples were transferred to cuvettes and were measured at A_{540 nm} on a Spectronic Biomate 3 (Thermo Fisher Scientific).

2.7.7 *E. coli* threonyl-tRNA synthetase (ThrRS) obafluorin (1)-binding assay

His₆-ThrRS, or its C480V variant, was first exchanged with MES buffer (pH 6.0) using an Amicon column (30 kDa MWCO). Reactions were performed in 100 µL volumes and were initiated by the addition of 95 µL of 10.5 µM His₆-ThrRS (10 µM final conc.)

to 5 μL of 2 mM obafluorin (**1**) dissolved in 100% MeCN (0.1 mM final conc.). 100% MeCN was used as a negative control and reactions were incubated at 28°C for 15 min, before analysis by UPLC-HRMS.

2.8 Metal-binding assays

2.8.1. Chrome azurol S (CAS) assay

500 μL volumes of CAS assay solution and 25 μM **1** (dissolved in 100% MeCN) were mixed in an Eppendorf tube. The CAS assay solution contains an iron-dye complex, which is disrupted by the presence of a strong iron-binding ligand, releasing the dye to produce a distinct colour change from blue through pink to orange (Schwyn and Neilands, 1987; Alexander and Zuberer, 1991).

2.8.2 4-(2-pyridylazo)-resorcinol (PAR) assay

1 μL 10 mM ZnSO_4 was added to a cuvette with 98 μL PAR assay solution. A serial dilution of **1** (dissolved in 100% MeCN) was performed and 1 μL was added to ZnSO_4 /PAR to give final concentrations of **1** of 1-50 μM . Solutions were mixed by pipetting before measuring spectra between 200-900 nm with a Lambda 35 UV/Visible Spectrophotometer (PerkinElmer). 1 μL 100% MeCN was applied as a negative control. The $\text{Zn}(\text{PAR})_2$ complex has a characteristic absorbance at 493 nm which is lost when ligands are introduced which compete for zinc-binding (Bandara *et al.*, 2009).

2.9 Antibacterial activity assays

2.9.1 Broth microdilution assay

MIC values were determined by broth microdilution in 96-well microtitre plates in duplicate, following approved standards of the Clinical and Laboratory Standards Institute and European Committee on Antimicrobial Susceptibility Testing (Kahlmeter *et al.*, 2006; Wiegand *et al.*, 2008), with some modifications. A stock solution of 2000 $\mu\text{g}/\text{mL}$ **1** was prepared in MeCN diluted 1:10 in MH broth and a dilution series was prepared down to 0.125 $\mu\text{g}/\text{mL}$. 50 μL of each dilution was then added to appropriate test wells (final conc. 0.0625-1000 $\mu\text{g}/\text{mL}$). Positive control wells contained 50 μL of 200 $\mu\text{g}/\text{mL}$ carbenicillin or 100 $\mu\text{g}/\text{mL}$ kanamycin (100 and 50 $\mu\text{g}/\text{mL}$ final concentrations, respectively). Negative control and sterility control wells contained 50 μL and 100 μL of sterile MH broth, respectively. Test strains were grown O/N in 5 mL LB cultures with appropriate selection at 37°C with 250 rpm. Cultures were standardised by dilution with LB against 0.5 McFarland standard solution to $\sim 1.5 \times$

10^8 CFU/mL, before subsequent dilution by 1:150 to 1×10^6 CFU/mL. 50 μ L of cell culture was then added to all microtitre plate wells (except for the sterility control) to a final inoculum concentration of 5×10^5 CFU/mL in 100 μ L reaction volumes. Microtitre plates were incubated at 37°C for ~16-18 h. The lowest concentration that showed no bacterial growth (no colour change or no turbidity) was recorded as the MIC.

2.9.2 Disc and agar diffusion assays

Test strains were grown O/N in 5 mL LB cultures containing appropriate selection. 50 μ L of each culture was used to inoculate 50 mL LB cultures, which were incubated at 37°C with 250 rpm until $OD_{600} \sim 0.4-0.5$. Cultures were used to inoculate molten SNA at 1% the final volume, before pouring into appropriately sized petri dishes to set. **1** was serially diluted from 2000-0.015 μ g/mL in 100% MeCN. 5 μ L of each dilution of **1** was spotted onto 5 mm Whatman filter discs (for disc diffusion method) or directly onto the SNA surface (for agar diffusion method). Kanamycin (50 μ g/mL final conc.) and carbenicillin (100 μ g/mL final conc.) were used as positive controls, and 100% MeCN was applied as a negative control. Plates were incubated at 37°C for ~16-18 h. The minimum inhibitory concentration (MIC) was defined as the lowest concentration of **1** that inhibited visible growth of the tested strains.

2.10 Chemical methods

2.10.1 Chemical synthesis and isolation of substrates

2.10.1.1 Obafluorin (**1**)

For the isolation of **1**, the supernatant from 6 L of culture broth was subjected to liquid-liquid partition using an equal volume of ethyl acetate. The organic fraction was concentrated under reduced pressure and the residue dissolved in MeCN. This was subjected to preparative reversed-phase HPLC using a Gemini® NX-C18 3u 110 Å 150 mm \times 4.6 mm column (Phenomenex); A: water; B: MeCN; gradient 0-5 min 10% B (v/v), 5-35 min, 10-100% B (v/v), 35-40 min 100% B (v/v), 40-41 min 100-10% B (v/v), 41-45 min, 10% B (v/v); flow rate was 20 ml/min; monitored at 254 nm). **1**: Pale yellow solid, (13 mg). ^1H NMR (400 MHz, MeCN- d_3) δ 8.23 (1H, d, $J = 8.8$ Hz), 8.10 (2H, d, $J = 8.8$ Hz), 7.45 (2H, d, $J = 8.8$ Hz), 7.20 (1H, dd, $J_1 = 8.2$ Hz, $J_2 = 1.27$ Hz), 7.04 (1H, dd, $J_1 = 7.9$ Hz, $J_2 = 1.2$ Hz), 6.82 (1H, t, $J = 7.96$ Hz), 5.75 (1H, dd, $J_1 = 8.5$ Hz, $J_2 = 6.2$ Hz), 5.05 (1H, m), 3.38 (1H, dd, $J_1 = 15.1$ Hz, $J_2 = 5.1$ Hz), 3.21 (1H, dd, $J_1 = 15.1$ Hz, $J_2 = 5.1$ Hz); ^{13}C NMR (100 MHz, MeCN- d_3) δ 171.47, 169.11, 150.43, 148.08, 147.02, 145.41, 131.20, 124.64, 120.50, 120.07, 118.68, 114.86, 78.50, 59.87, 36.16 ppm HRMS (m/z): $[\text{M}+\text{H}]^+$ calculated for $\text{C}_{17}\text{H}_{14}\text{N}_2\text{O}_7$, 359.0874,

found, 359.0872, Diff. (ppm) = -0.56. The NMR data is consistent with published values (Pu *et al.*, 1994).

2.10.1.2 Ethyl 5-(4-nitrobenzyl)-4,5-dihydrooxazole-4-carboxylate

Ethyl 5-(4-nitrobenzyl)-4,5-dihydrooxazole-4-carboxylate was synthesized according to a literature procedure (Rao *et al.*, 1991) using **11** (990 mg, 6 mmol) and ethyl isocynoacetate (750 mg, 6.6 mmol) to yield 910 mg of the racemic *trans*-diastereoisomer. Yield: 55%. ¹H NMR (400 MHz, CDCl₃) δ 8.08 (2H, d, *J* = 8.8 Hz), 7.36 (2H, d, *J* = 8.8 Hz), 6.87 (1H, d, *J* = 9.4 Hz), 4.72 (1H, dd, *J*₁ = 9.4 Hz, *J*₂ = 2.0 Hz), 4.41-4.36 (1H, m), 4.22-4.13 (2H, m), 2.92-2.78 (2H, m), 1.23 (3H, t, *J* = 7.2 Hz); ¹³C NMR (100 MHz, CDCl₃) δ 170.1, 161.8, 146.8, 145.3, 130.3, 123.6, 72.1, 62.2, 54.6, 40.1, 14.0 ppm; HRMS (*m/z*): [M+H]⁺ calculated for: C₁₃H₁₅N₂O₅, 279.0975; found, 279.0978, Diff. (ppm) = 1.07.

2.10.1.3 (2*SR*,3*RS*)-2-amino-3-hydroxy-4-(4'-nitrophenyl)butanoic acid (**AHNB**; **2**)

2 was synthesized according to a literature procedure (Rao *et al.*, 1991). Starting from 720 mg of ethyl 5-(4-nitrobenzyl)-4,5-dihydrooxazole-4-carboxylate we obtained 570 mg of **2**. Yield: 92%. ¹H NMR (400 MHz, D₂O) δ 8.31 (2H, d, *J* = 8.5 Hz), 7.63 (2H, d, *J* = 8.5 Hz), 4.44 (1H, m), 3.84 (1H, m), 3.27 (1H, dd, *J*₁ = 14.0 Hz, *J*₂ = 3.2 Hz), 3.05 (1H, dd, *J*₁ = 14.0 Hz, *J*₂ = 1.6 Hz); ¹³C NMR (100 MHz, D₂O) δ 172.5, 146.6, 145.9, 130.5, 123.9, 70.3, 59.2, 39.8 ppm HRMS (*m/z*): [M+H]⁺ calculated for: C₁₀H₁₃N₂O₅, 241.0819; found: 241.0801, Diff. (ppm) = -7.47. NMR data are in agreement with published values for the (2*S*,3*R*)-enantiomer.

The enzymatic synthesis of **2** was achieved by scaling up (to 2 ml) the analytical conditions for the ObaG discontinuous assay described above (2.7.1). The reaction was quenched after 2 h by the addition of MeOH (2 ml) and the solution was concentrated under reduced pressure. The resulting crude product was repeatedly subjected to solid-phase extraction for further purification and to remove excess buffer. Elution with 25% MeOH yielded **2** (1.8 mg, 38%). NMR data were consistent with those of the synthetic reference standard. [α]_D + 48°, (c = 0.18, H₂O) (Pu *et al.*, 1994; [α]_D + 50°, (c = 0.18, H₂O)).

2.10.1.4 (2*SR*,3*RS*)-2-amino-3-hydroxy-4-(4'-aminophenyl)butanoic acid (AHAB; **9**)

9 was synthesized according to a literature procedure (Herbert *et al.*, 1994). Starting from 25 mg of **2** we obtained 7 mg of **10** as a yellow solid. ¹H NMR (400MHz, D₂O) δ 7.37 (2H, d, *J* = 8.4 Hz), 7.28 (2H, d, *J* = 8.1 Hz), 4.39 (m, 1H), 4.05 (1H, d, *J* = 4.0 Hz), 3.01 (1H, dd, *J*₁ = 14.1 Hz, *J*₂ = 4.3 Hz), 2.82 (1H, dd, *J*₁ = 9.9 Hz, *J*₂ = 14.2 Hz); ¹³C NMR (100MHz, D₂O) δ 170.35, 138.28, 130.86, 128.44, 123.17, 69.71, 57.16, 38.77 ppm HRMS (m/z): [M+H]⁺ calculated for: C₁₀H₁₅N₂O₃, 211.1077 [M+H]⁺; found, 211.1082, Diff. (ppm) = -2.37.

2.10.2 Analytical methods and instrumentation

I performed HR-LCMS and HPLC analysis of extracts from mutational analysis and biochemical experiments. Dr. Daniel Heine performed His₆-ObaG transaldolase assay NMR experiments, isolated enzymatically synthesised compounds and synthesised substrates. Prof. Barrie Wilkinson also helped synthesise substrates. Dr. Daniel Heine and Dr. Gerhard Saalbach performed analyses of ThrRS 1-binding experiments. Dr. Zhiwei Qin Isolated **1** and carried out NMR analysis.

NMR spectra were recorded on a Bruker AVANCE III 400 MHz spectrometer. Chemical shifts are reported in parts per million (ppm) relative to the solvent residual peak of chloroform-d₁ (¹H: 7.24 ppm, singlet; ¹³C: 77.00 ppm, triplet), methanol-d₄ (¹H: 3.30 ppm, quintet; ¹³C: 49.00 ppm, septet) or water-d₂ (¹H: 4.79 ppm).

Semi-preparative and analytical HPLC was performed using an 1100 system (Agilent Technologies). Extracted samples from mutagenesis and complementation experiments were analysed using a Gemini® NX-C18 3u 110 Å, 150 mm × 4.6 mm column (Phenomenex) with a gradient elution: MeCN/0.1% (v/v) TFA (H₂O) - gradient from 10/90 to 100/0 0–15 min, 100/0 for 15–16 min, gradient to 10/90 16–16.50 min and 10/90 for 16.50–23 min; flow rate 1 ml/min; injection volume 10 μL.

Biochemical assays were analysed using a Synergi 4 mm Fusion-RP 80 Å LC column 250 × 4.6 mm (Phenomenex) with a gradient elution: MeOH/0.1% (v/v) TFA (H₂O) - gradient from 10/90 to 100/0 0–14 min, 100/0 for 14–18 min, gradient to 10/90 18–18.50 min and 10/90 for 18.50–23 min; flow rate 1 ml/min; injection volume 5 μL (15 μL for the reverse reaction experiments).

UPLC-MS measurements were performed on a Nexera X2 liquid chromatograph (LC-30AD) LCMS system (Shimadzu) connected to an autosampler (SIL-30AC), a Prominence column oven (CTO-20AC) and a Prominence photo diode array detector (SPD-M20A). A Kinetex 2.6 mm C18 100 Å, 100 2.1 mm column (Phenomenex) was used for LCMS. The UPLC-System was connected to a LCMS-IT-TOF Liquid Chromatograph mass spectrometer (Shimadzu).

UPLC-HRMS data for ThrRS 1-binding experiments were acquired with an Acquity UPLC system connected to a Synapt G2-Si high resolution mass spectrometer (Waters). Analytical UPLC was performed using an Aeris widepore 3.6 µm C4, 50 × 2.1 mm column (Phenomenex) with a gradient elution: 0.1% (v/v) formic acid (MeCN)/0.1% (v/v) formic acid (H₂O) - 95/5 for 0-1 min, gradient to 10/90 2-6 min, 10/90 for 6-7 min, gradient to 90/10 7-7.1 min, 90/10 for 7.1-9 min, gradient to 10/90 for 9-14 min, gradient to 95/5 for 14-14.1 min, 95/5 for 14.1-16 min; flow rate 400 µL/min; injection volume 5 µL. MS spectra were acquired with a scan time of 1 s in the range of $m/z = 100 - 2000$ in positive MS resolution mode. The following parameters were used: capillary voltage of 3 kV, sampling Cone 40 V, source offset: 80 V, source temperature of 120 °C, desolvation temperature of 350 °C, cone gas flow of 10 L/h, desolvation gas flow of 900 L/h and a nebulizer pressure of 6 bar. A solution of sodium formate was used for calibration and a solution of leucine enkephalin (H₂O/MeOH/formic acid: 49.95/49.95/0.1) was used as lock mass, which was injected every 15 s. The lock mass was acquired with a scan time of 0.3 s and 3 scans to average. The lock mass (556.2766) was applied during data acquisition.

ThrRS 1-binding assays were further analysed following tryptic digest. Reaction mixtures were first denatured with sodium deoxycholate, treated with DTT and digested with trypsin according to standard procedures. For LC-MS/MS analysis of peptide fragments from, an Orbitrap-Fusion™ mass spectrometer (Thermo Fisher Scientific) equipped with an UltiMate™ 3000 RSLCnano System using an Acclaim PepMap C18 column (2 µm, 75 µm x 500 mm, Thermo Fisher Scientific) was used. Aliquots of the tryptic digests were loaded and trapped using a pre-column which was then switched in-line to the analytical column for separation. Peptides were eluted with a gradient of 5-40% MeCN in ddH₂O/0.1% HCOOH at a rate of 0.5%/min. The column was connected to a 10 µm SilicaTip™ nanospray emitter (New Objective) for infusion into the mass spectrometer. Data dependent analysis was performed using parallel HCD/CID fragmentation using a top20 method with following parameters: positive ion mode, orbitrap MS resolution = 120k, mass range (quadrupole) = 300-

1800 m/z , MS/MS isolation window 1.6 Da, threshold 2.5e4, AGC target 1e4, max. injection time 35 ms, dynamic exclusion 2 counts, 30 s exclusion, exclusion mass window ± 7 ppm. MS scans were saved in profile mode while MS/MS scans were saved in centroid mode. Raw files from the Orbitrap were processed with MaxQuant (version 1.6.0.1) to generate re-calibrated peaklist-files which were used for database searches using an in-house Mascot Server (version 2.4). The searches were performed on a custom database containing the sequence of interest and on a common contaminants database using trypsin/P with 2 missed cleavages, carbamidomethylation (C), oxidation (M), acetylation (protein N-terminus), and 1 (Cysteine, +358.0801 Da) as variable modifications. Mass tolerances were 6 ppm for precursor ions and 0.6 Da for fragment ions. Mascot search results were imported and evaluated in Scaffold 4.4.1.1 (proteomsoftware.com) with thresholds of 99% and 95% for proteins and peptides.

Solid phase extraction was carried out using Discovery DSC-18 SPE Tubes, filled with 1,000 mg of octadecyl-modified, endcapped silica gel (Supelco). The specific optical rotation of compounds was measured with a Model 341 Polarimeter (PerkinElmer, Inc.).

2.11 Phylogenetic analysis

Amino acid sequences of L-TAs and SHMTs were obtained from a previous phylogenetic study (Liu *et al.*, 2015) and were combined with additional L-TA and SHMT sequences for enzymes which have been characterized or described in the literature, and were obtained from the National Centre for Biotechnology Information (NCBI) GenBank database (Clark *et al.*, 2016) and the Protein Data Bank (PDB) (Berman *et al.*, 2000). Blastp (Basic Local Alignment Search Tool – Altschul *et al.*, 1990) searches for L-TAs or SHMTs involved in NP biosynthesis were performed to identify enzymes from these families associated with specialized metabolism. A selection of 10 amino acid sequences for 3-amino-5-hydroxybenzoic acid synthases (AHBASs) described in the literature were obtained to function as an outgroup for phylogenetic analysis as they have been shown to share a recent common ancestor of L-TAs and SHMTs (Contestabile *et al.*, 2001) (Source organisms and amino acid sequence GenBank accession numbers used are reported in Appendix 1 - Supplementary Table 2). Sequences were initially aligned (all alignments performed with default settings) using ClustalX2 (Larkin *et al.*, 2007), before manual trimming of sequences at N- and C-termini to remove aberrant sequences (e.g. histidine tags)

that might interfere with the alignment. Trimmed sequences for the phylogeny illustrated in Figure 6.1 are reported in Supplementary Sequence Data 1 (Appendix 2). Several iterations of alignment and trimming were repeated with different degrees of trimming to ensure that the final tree was relatively consistent and robust. MUSCLE (Edgar, 2004), in addition to ClustalX2, was trialled for initial alignment before trimming. Trimmed sequences were finally re-aligned with T-Coffee (Notredame *et al.*, 2000), and phylogenetic tree inference was performed using maximum likelihood/rapid bootstrapping under the Generalised Time-Reversible (GTR) model using RAxML-HPC BlackBox (8.2.8) (Stamatakis, 2014) via the CIPRES Science Gateway portal (Miller *et al.*, 2010). The JTT Protein Substitution Matrix was used and all other parameters were set to default values.

2.12 Software

Software used for *in silico* analyses are listed in Table 2.7

Software	Use	Reference
antiSMASH (v2.0-4.0) (https://antismash.secondarymetabolites.org)	Identification of putative BGCs in DNA sequence data	Blin <i>et al.</i> , 2017
Artemis (v16.0.0) (http://www.sanger.ac.uk/science/tools/artemis)	Genome sequence data visualisation	Rutherford <i>et al.</i> , 2000
BioEdit (v7.1.11)	Amino acid sequence alignment and trimming	Hall, 1999
BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi)	Homology searches of DNA and protein sequences	Altschul <i>et al.</i> , 1990
BOXSHADE (v3.21) (http://www.ch.embnet.org/software/BOX_form.html)	For creating sequence alignment outputs	N/A
ChemDraw (v16.0)	Property prediction of chemicals and for generating illustrations	Cambridgesoft
Cipres (v3.3) (https://www.phylo.org/)	Phylogenetic tree inference	Miller <i>et al.</i> , 2010
ClustalX2 (v2.0)	Amino acid sequence alignment and trimming	Larkin <i>et al.</i> , 2007
DNAPlotter (v1.11)	Genome visualisation	Carver <i>et al.</i> , 2009
ExpASy Compute pI/Mw Tool (http://web.expasy.org/compute_pi/)	Calculate protein pI and Mw	Gasteiger <i>et al.</i> , 2005

ExpASY Translate Tool (http://web.expasy.org/translate/)	Translation of nucleotide sequences	Gasteiger <i>et al.</i> , 2005
FigTree (v4.1.3) (http://tree.bio.ed.ac.uk/software/figtree/)	Phylogenetic tree visualisation	N/A
GraphPad Prism (v5.04) (https://www.graphpad.com/scientific-software/prism/)	Biochemical assay and kinetic data analysis	GraphPad Software Inc.
MaxQuant (v1.6.0.1) (http://maxquant.org)	Processing of Orbitrap raw data files	Cox and Mann, 2008
Mascot Server (v2.4) (http://www.matrixscience.com/)	Analysis of protein MS data	N/A
MUSCLE (v3.8.31) (http://www.drive5.com/)	Multiple sequence alignment	Edgar, 2004
NEBcutter® (v2.0) (http://nc2.neb.com/NEBcutter2/)	Identification of restriction sites in DNA sequence data	New England Biolabs
NEBuilder® (https://nebuilder.neb.com/)	Design of primers for Gibson Assembly	New England Biolabs
Pfam (v30.0-31.0) (http://pfam.xfam.org/)	Protein domain annotation	Finn <i>et al.</i> , 2016
Phyre2 (v2.0) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)	Protein folding recognition	Kelley <i>et al.</i> , 2015
Prodigal (v2.0) (http://prodigal.ornl.gov/)	PCS/gene prediction	Hyatt <i>et al.</i> , 2010
Rast (v2.0) (http://rast.nmpdr.org/)	Prokaryotic genome annotation	Overbeek <i>et al.</i> , 2014
Scaffold (v4.4.1.1) (proteomsoftware.com)	Evaluation of Mascot search results	N/A
T-Coffee (v11.00.8) (http://www.tcoffee.org/Projects/tcoffee/)	Multiple sequence alignment	Notredame <i>et al.</i> , 2000
The Sequence Manipulation Suite (https://www.bioinformatics.org/sms/index.html)	Reverse complementation of DNA sequence data	Stothard, 2000
Vector NTI Advance (v11.5.2) (https://www.thermofisher.com/us/en/home/life-science/cloning/vector-nti-software/vector-nti-advance-software.html)	Vector map construction and visualisation, and visualisation and contig assembly of sequence data (ContigExpress)	Thermo Fischer Scientific
WebACT (v13.0.0) (http://www.webact.org/WebACT/home)	Genome pair-wise comparison and visualisation	Carver <i>et al.</i> , 2005; Abbot <i>et al.</i> , 2007

Table 2.7. Software used for in silico analyses in this work.

Chapter 3:
Identification and delineation
of the obafluorin (**1**) BGC

Chapter 3: Identification and delineation of the obafluorin (1) BGC

3.1 Introduction

In addition to the discovery of novel NP scaffolds, access to entire genomes has allowed NP researchers to revisit a wealth of previously published molecules. Their structure and biosynthesis had only been characterised using chemistry-based approaches, but with advances in molecular approaches and genome sequencing, their biosynthesis can now be dissected at the genetic and biochemical levels. Of particular interest are NPs that possess unusual structural features which might endow a novel mode of action, and/or allow the discovery of unusual biosynthetic enzymes for synthetic applications. The structure of a NP and its breakdown products/co-metabolites allow a retrobiosynthetic approach to be used to predict the likely biosynthetic steps in its pathway. In the case of the biosynthesis of obafluorin (1), additional data are available from stable isotope labelling experiments which were described in detail in Chapter 1 (Herbert and Knaggs, 1988; 1990; 1992a; 1992b). Based on their results, Herbert and Knaggs identified 4-APP (7) or 4-NPA (8) and glyoxylate as the likely precursors of the unusual β -OH- α -AA AHNB (2) (Figure 3.1), which when condensed with 2,3-DHBA (6), form a *pseudo*-dipeptide precursor to 1. This proposal for the biosynthesis of 1 provided the starting point for this work.

In this Chapter, the sequencing and *in silico* analysis of the *P. fluorescens* ATCC 39502 genome are described. Comparative genomics approaches were first applied to highlight any unusual features in the ATCC 39502 genome. Putative BGCs were then investigated in order to identify the likely BGC for 1 based on anticipated homology to previously characterised biosynthetic pathways and data from previous feeding studies. Biosynthetic predictions are made about the timing and functions of enzymes in the pathway.

3.2 Sequencing and annotation of the *P. fluorescens* ATCC 39502 genome

A PCR was performed with degenerate primers F27 and R1492 (Weisburg *et al.*, 1991; Heuer *et al.*, 1998) to amplify the 16S rRNA gene from the ATCC 39502 strain. This was sequenced and shared 99% query cover and identity at the nucleotide level to *P. fluorescens* SBW25. High molecular weight genomic DNA was subsequently prepared using a salting out method (Materials and methods section 2.4.1) and

7 and 9, X = H
8 and 2, X = O

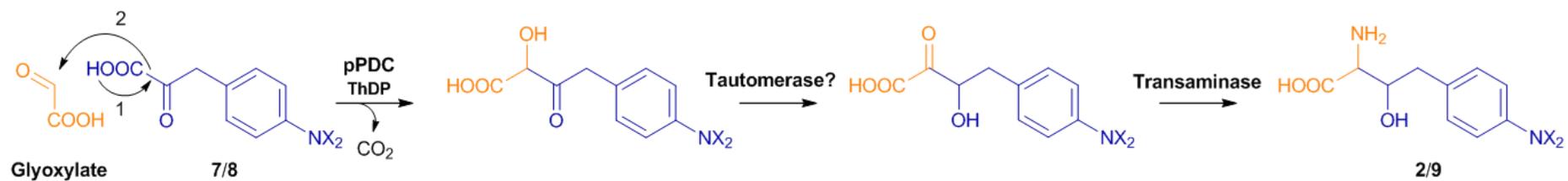


Figure 3.1. Biosynthesis of the β -OH- α -AA acid AHNB (2) as proposed by Herbert and Knaggs (1990).

submitted to The Earlham Institute (formerly The Genome Analysis Centre) for sequencing using SMRT technology on the Pacific Biosciences (PacBio) RSII platform, and assembly using the HGAP2 pipeline. Given the highly repetitive nature of modular NRPSs and PKSs, shorter sequence read lengths can result in assembly issues. PacBio was therefore preferred over other sequencing technologies (Gomez-Escribano *et al.*, 2016) as it is capable of achieving average read lengths in excess of 20 kb (<http://www.pacb.com/smrt-science/smrt-sequencing/read-lengths/>). In comparison, Illumina's HiSeq typically generates average read lengths of ~500 bp. In total, 110,545 sequence reads (mean read length, 5,700 bp) were used to assemble a single contig circular genome of just over 6.15 Mbp (GC content, 60.3%) with 115x coverage (80x required for 99.99% accuracy) (Figure 3.2). This contig was submitted to the Prodigal server (Hyatt *et al.*, 2010) for identification of protein coding sequences (PCSs) and initial putative gene functions were assigned using the RAST server (Overbeek *et al.*, 2014).

3.3 Preliminary analysis of the *P. fluorescens* ATCC 39502 genome

P. fluorescens spp. are known to have highly variable genomes, to the extent that they have been re-branded as a complex rather than as an individual species (Silby *et al.*, 2009; 2011). The ATCC 39502 genome comprised specific marker genes identifying it as belonging to the *P. fluorescens* phylogroup of the *P. fluorescens* complex (Garrido-Sanz *et al.*, 2017). Comparative genomic analyses were performed to compare the ATCC 39502 sequenced genome against those of three other fully-sequenced model *P. fluorescens* strains to identify any unusual features. A basic comparison of general genomic characteristics revealed that all four strains are relatively similar: 6-7 Mbp in size, 87-89% coding density, 60-61% GC content and comparable numbers of tRNA and rRNA genes (Table 3.1).

3431 PCSs are conserved between all four strains, representing a 'core' genome comprising 56-65% of total PCSs in each genome: this is in accordance with a similar comparison of *P. fluorescens* genomes conducted previously (Silby *et al.*, 2009). In order to identify the location of 'core' PCSs within the genome, a pairwise tblastx comparison was performed by Dr. Govind Chandra, comparing the SBW25, Pf0-1 and F113 genomes against that of ATCC 39502. Comparisons were visualised using the Act Comparison Tool (Carver *et al.*, 2005) which revealed that ATCC 39502 shares greatest synteny with SBW25 (Figure 3.3), and that the majority of homologous PCSs shared between strains are located close to the origins of replication in each strain, again recapitulating previous observations (Silby *et al.*,

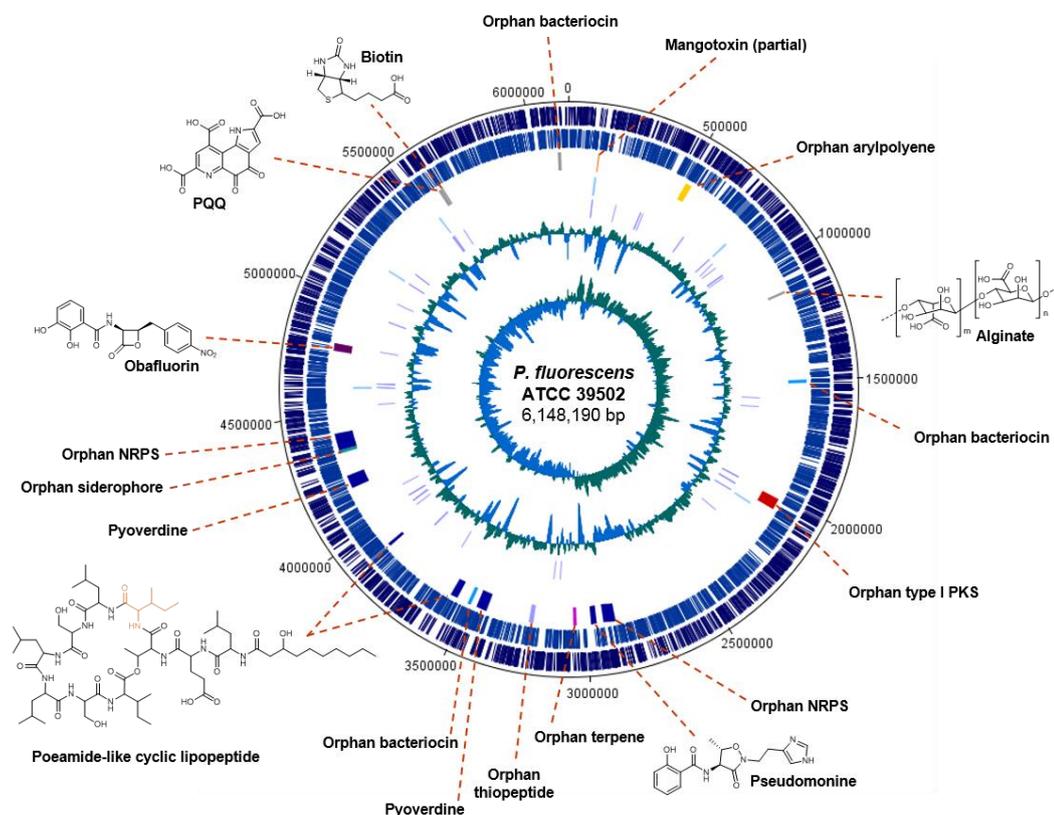


Figure 3.2. Illustration of the *P. fluorescens* ATCC 39502 genome. The outer scale designates coordinates in base pairs (bp), with the origin of replication located at 0 bp. The first (outermost) track shows predicted PCSs on the + strand and the second shows predicted PCSs on the - strand. The third track comprises BGCs predicted by antiSMASH 4.0 (Blin et al., 2017), as described in the main text. The fourth track shows tRNA genes and the fifth track shows rRNA genes. The sixth track is a representation of GC plot and the seventh (innermost) track represents GC skew. Structures of NPs encoded in the genome are illustrated and linked to their respective BGCs. The predicted structure of a poamide-like CLP has been illustrated with the predicted isoleucine (highlighted in orange), rather than the leucine residue present in poamide.

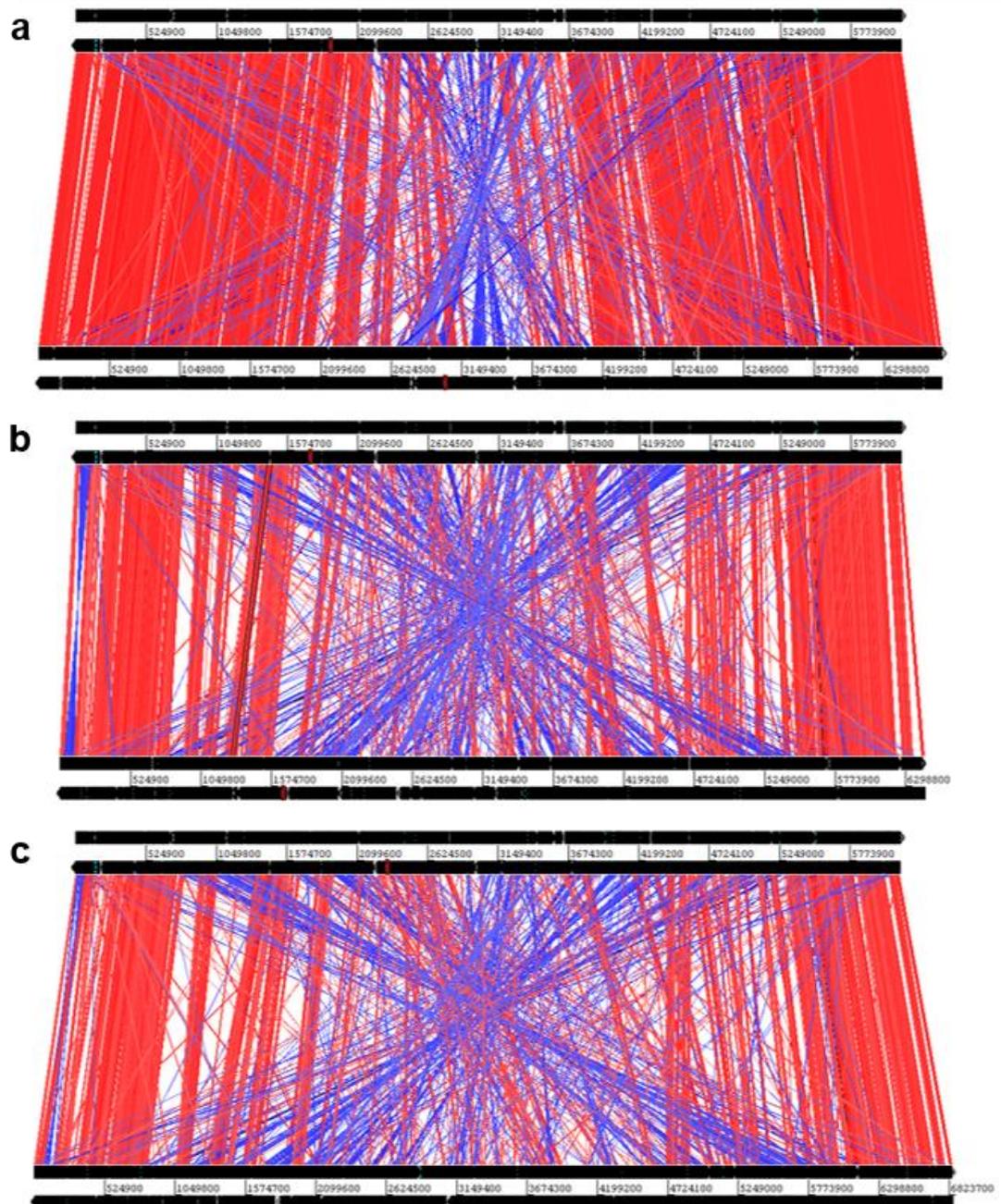


Figure 3.3. Assessment of genome synteny between *P. fluorescens* ATCC 39502 and three fully-sequenced model *P. fluorescens* strains. The top genome in each of the three panels represents ATCC 39502 and the bottom genome belongs to (a) SBW25; (b) Pf0-1; and (c) F113. *P. fluorescens* genomes are circular and the origins of replication for each strain are located at the first bp in the genome (i.e. at the start and end of the genomes represented here). Red bars between genomes represent individual *tblastx* matches and inverted matches are blue.

2009). The SBW25, Pf0-1 and F113 genomes were also submitted to RAST, which allocates general functions to PCSs. Comparison of allocated gene functions across the four genomes revealed that they share an extremely similar distribution in the function of their PCSs (Figure 3.4), with amino acid, protein and carbohydrate metabolism predominating, and high numbers of PCSs dedicated to membrane transport, iron acquisition and metabolism, and stress response, perhaps reflecting the ability of this species to adapt to a broad range of environments. ATCC 39502 thus displays typical, characteristic genomic features of members of the *P. fluorescens* complex.

General characteristics of <i>P. fluorescens</i> genomes				
<i>P. fluorescens</i> strain	ATCC 39502	SBW25	Pf0-1	F113
Number of bp	6,148,190	6,722,539	6,438,405	6,845,832
Number of PCSs	5,336	5,976	5,707	6,055
Coding %	87.3	87.3	88.5	87.1
%GC	60.3	60.5	60.5	60.8
tRNA genes	69	66	73	66
rRNA genes	19	16	19	16
Predicted BGCs* (CF)	17 (49)	9 (54)	10 (45)	10 (54)

Table 3.1. Comparison of *P. fluorescens* ATCC 39502 genomic characteristics with those of other sequenced model *P. fluorescens* strains. * refers to the number of putative BGCs identified by antiSMASH 4.0 (Blin *et al.*, 2017), and the number identified when the ClusterFinder algorithm was applied is given in brackets.

3.4 Identifying BGCs in the *P. fluorescens* ATCC 39502 genome

The ATCC 39502 genome sequence was submitted to antiSMASH 4.0 (Blin *et al.*, 2017) and, using the ClusterFinder algorithm (Cimermancic *et al.*, 2014), 49 putative gene clusters were identified, with 17 assigned to a specific NP class (Table 3.2). This is above average compared to other *P. fluorescens* strains in which nine or ten BGCs were identified by antiSMASH (Table 3.1). Each putative BGC was analysed manually using BLAST (Altschul *et al.*, 1990) and Pfam (Finn *et al.*, 2016) searches to determine their likely products. All ATCC 39502 BGCs assigned to a specific NP class or product are illustrated in Figure 3.2, along with their genomic location. Putative BGCs are concentrated in the more variable region of the genome, away from the origin of replication (position 0) and the 'core' PCSs, consistent with the variation in

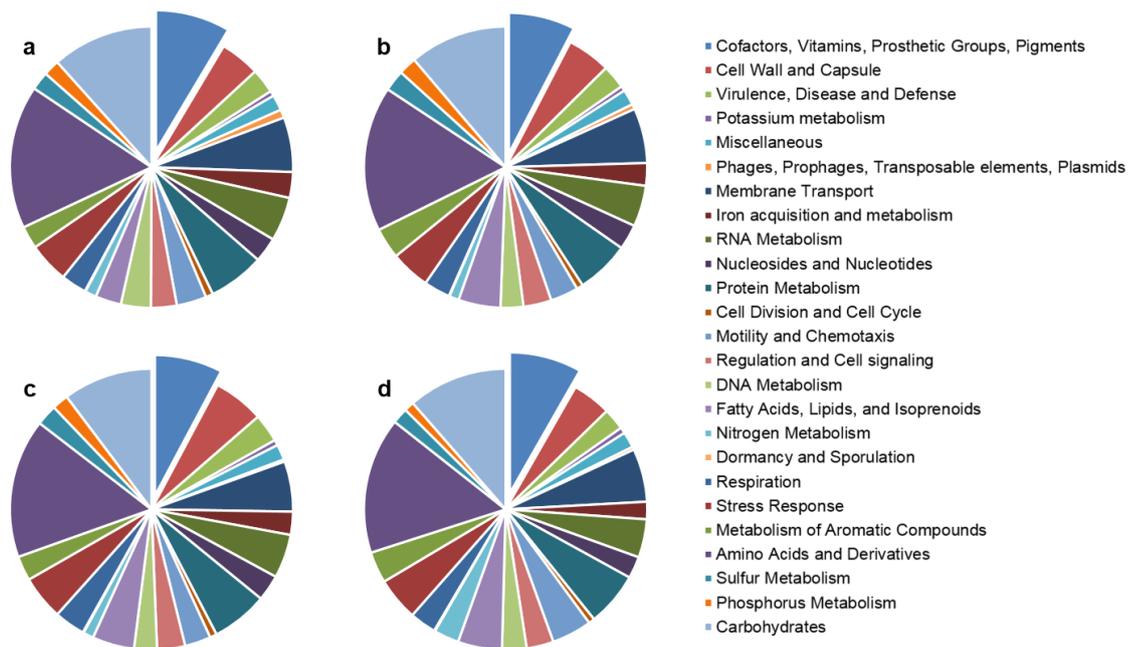


Figure 3.4. Functional assignment of PCSs in the genomes of four fully-sequenced *P. fluorescens* strains by the RAST server. (a) ATCC 39502 (52%); (b) SBW25 (53%); (c) Pf0-1 (53%); and (d) F113 (52%). Percentages represent the PCSs identified by RAST that could be assigned a function. Pie chart segments reflect the relative number of those PCSs assigned to each functional group. The ‘cofactors, vitamins, prosthetic groups, pigments’ segment has been pulled out as a point of reference, with subsequent categories in the key occurring in a clockwise direction.

BGC composition observed between *P. fluorescens* strains and the common transmission of BGCs via MGEs.

antiSMASH Cluster No.	BGC type	Position		BGCs of compounds with highest similarity (%)
		From	To	
1	Other	108144	151521	Mangotoxin (57%)
2	Arylpolyene	488265	531840	APE_Vf (40%)
3	Bacteriocin	1500434	1511309	-
4	T1pks-Resorcinol	2009855	2059104	-
5	Nrps	2879189	2927378	-
6	Nrps	2939150	2990150	Pseudomonine (100%)
7	Terpene	3042675	3064894	-
8	Thiopeptide	3209903	3235479	Lipopolysaccheride (5%)
9	Nrps	3400040	3465917	Tawaichelin (11%)
10	Bacteriocin	3491452	3502288	-
11	Nrps	3531661	3599079	Poeamide (71%)
12	Nrps	3886949	3933221	Poeamide (100%)
13	Nrps	4189378	4242289	Pyoverdine (9%)
14	Siderophore	4339809	4351734	-
15	Cf_fatty_acid-Nrps	4354109	4424976	WS9326 (7%)
16	Homoserine lactone-NRPS	4762163	4815888	Lipopeptide_8D1-1 (6%)
17	Bacteriocin	6104950	6115795	-

Table 3.2. Summary of *P. fluorescens* ATCC 39502 antiSMASH output.

Surprisingly, only two BGCs contained elements of PK biosynthesis, despite the large number of known pseudomonad PKS and hybrid NRPS-PKS biosynthetic pathways (Gross and Loper, 2009). One was a type I PKS with no homology to any known BGCs, and the other was a putative arylpolyene-encoding BGC. Arylpolyenes are pigmented compounds that have been recently recognised as being broadly distributed among the proteobacteria, in which they are proposed to play important roles in mitigating oxidative stress through their ability to effectively quench reactive oxygen species (Cimermancic *et al.*, 2014; Schöner *et al.*, 2016). Other assigned BGCs include a single thiopeptide cluster and a siderophore biosynthesis cluster, the latter comprising an *iucA/iucC*-like siderophore synthetase gene similar to that involved in aerobactin biosynthesis (De Lorenzo *et al.*, 1986). Both of these BGCs have numerous homologues in other *Pseudomonas* spp., but neither have been previously characterised. A saccharide cluster was found to share 37% homology to the pseudopyronine BGC (Bauer *et al.*, 2015), however no pseudopyronine synthase gene was found in the ATCC 39502 genome, indicating this cluster is probably inactive, or may perform an alternate role.

So far neither terpenoid metabolites, nor terpene synthases, have been reported from *Pseudomonas* spp. in the literature. Terpenes represent one of the largest NP classes and encompass the greatest chemical diversity of all NPs. They originate from the C₅ substrates dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP), where one or more IPP units are condensed with DMAPP in a head-to-tail fashion to form the basic terpenoid scaffolds geranyl diphosphate (GPP - C₁₀), farnesyl diphosphate (FPP - C₁₅), or geranylgeranyl diphosphate (GGPP - C₂₀) (Oldfield and Lin, 2012). These units can be further condensed together to form larger terpene scaffolds, before subsequent cyclisation and/or any number of other post-translational modifications (PTMs). Intriguingly, antiSMASH 4.0 identified a BGC comprised of 11 PCSs, including a dimodular terpene synthase/cytochrome P450 enzyme, IPP isomerase and class I SAM methyltransferase. Whilst the cluster has a number of homologues in other *Pseudomonas* spp., it could not be linked to any known terpene BGCs. A terpene-NRPS BGC is identified by antiSMASH in the genome of *P. fluorescens* Pf0-1 genome, but the NRPS element corresponds to homologues of *pvdG/I/J* in pyoverdine biosynthesis (described below) and the terpene element refers to a terpene synthase enzyme located just under 3 kb upstream, which does not appear to be involved in a specialised metabolite BGC context.

BLAST searches were performed to try and identify PCSs for other well-characterised specialised metabolites from *P. fluorescens* such as pyoluteorin, pyrrolnitrin, 2,4-DAPG and HCN (Gross and Loper, 2009), but no homologues could be found in *P. fluorescens* ATCC 39502. 32 of the putative BGCs predicted by the ClusterFinder algorithm were either classed as saccharides or were not assigned to a specific NP class and their likely functions, if any, were very difficult to determine. Amongst them, primary metabolite BGCs for biotin, pyrroloquinolone quinone (PQQ) and alginate biosynthesis were identified, in addition to loci encoding factors of polymyxin resistance and a cluster of flagellar-related PCSs (Figure 3.2).

By far the most abundant NP BGC class in *P. fluorescens* ATCC 39502 is the NRPs, with eight candidate clusters comprising NRPS-like genes. Among them, six showed homology to known NRP BGCs. The entire BGC for the siderophore pseudomonine (Sattely and Walsh, 2008) is present (100% of genes show similarity), in addition to two BGCs which appear to be involved in the biosynthesis of pyoverdine. The first comprises homologues of *pvdSGL*, with PvdL being the NRPS responsible for assembling scaffold for the characteristic tripeptide chromophore moiety (Böckmann *et al.*, 1997). The second BGC comprises two NRPSs, one of which is a homologue

of *pvdD*, the product of which is the terminal enzyme in the assembly of the variable peptidic chain in the pyoverdines. Based on the number of modules comprised by these two NRPSs, the peptidic side chain is likely to comprise seven amino acids. *pvd* loci for pyoverdine modification, regulation, export, siderophore detection and uptake are also found located elsewhere in the genome.

An incomplete cluster comprising *mgoABC* homologues (60-70% amino acid sequence similarity – *mgoD* missing) involved in the biosynthesis of the *P. syringae* antimetabolite mangotoxin is also present (Arrebola *et al.*, 2012). The *mboABCDEF* operon, which is only found in mangotoxin-producing strains and has been shown to be essential for production (Carrión *et al.*, 2012), is absent however, so the function of *mgo* homologues in the ATCC 39502 genome is unclear. Two unlinked clusters comprising three NRPSs between them share extremely high homology to CLP clusters of the orfamide family, with 80-90% shared amino acid identity with *poaABC* in the poeamide-producer *P. poea* RE*1-1-14 (Zachow *et al.*, 2015) – the key difference is that the A domain in module 4 of the *poaB* homologue in *P. fluorescens* ATCC 39502 appears to encode isoleucine specificity, as opposed to leucine.

Based on the findings described in section 3.1, predictions could be made about PCSs anticipated in a putative *oba* BGC. These include PCSs involved in the biosynthesis of 4-APA (**3**) and **6**, and one encoding a ThDP-dependent enzyme proposed by Herbert and Knaggs to catalyse a C-C bond forming reaction between glyoxylate and **7**(or **8** depending on when the nitro group is introduced) to generate the acyloin precursor to **2** (Figure 3.1 - Herbert and Knaggs, 1990). A transaminase should also be present in a putative **1** BGC for the introduction of the α -amino group on **2**. *Pseudomonas* spp. are known to produce many structurally diverse NP peptides (Gross and Loper, 2009), and given that many are derived from NRPSs, it is reasonable to hypothesise that **1** is similarly assembled using **2** and **6**. Based on the biosynthetic logic of these multienzyme complexes, a dimodular NRPS should be sufficient to catalyse peptide bond formation between **2** and **6**, with at least one A domain with substrate specificity for **6**. Given the nonproteinogenic nature of **2**, it would be unlikely that A domain substrate specificity predicting software would be able to identify the correct substrate for the second A domain in an *oba* BGC NRPS. Finally, an *N*-oxygenase similar to that observed in pyrrolnitrin biosynthesis (Lee *et al.*, 2005) is the most likely candidate for the introduction of the aromatic nitro group of **1**.

One designated NRPS BGC stood out in particular as being the likely candidate for the biosynthesis of **1** and is described in the subsequent sections.

3.5 *in silico* analysis of the obafluorin (1) BGC

The minimal *oba* BGC, is approximately 19 kbp in length. PCSs upstream of the cluster (*orf12345678*) could also play roles in precursor supply and export, potentially extending the cluster to 30 kbp (Figure 3.5). PCSs are reported in Table 3.3. Domains identified within each PCS by BLAST and Pfam are also stated. In the following sections, the likely roles of *oba* PCSs in the pathway are described.

Protein / Gene	Size (Da/bp)	Proposed function	Conserved domains (Based on Pfam and NCBI conserved domains)	GenBank accession ID of protein homologue (query cover/identity %)
Orf-6/ <i>orf-6</i>	50833.13/ 1428	Arginine/ornithine antiporter	Amino acid permease domain	WP_065949362.1 (100/100)
Orf-5/ <i>orf-5</i>	15267.84/ 1428	Arginine/ornithine antiporter	Amino acid permease domain	WP_065949361.1 (100/100)
Orf-4/ <i>orf-4</i>	46453.57/ 1257	Arginine deaminase	Amidino transferase family domain	WP_065949360.1 (100/99)
Orf-3/ <i>orf-3</i>	37760.61/ 1011	Ornithine carbamoyltransferase	Carbamoyl-P binding domain; aspartate/ornithine binding domain	WP_065949359.1 (100/99)
Orf-2/ <i>orf-2</i>	32651.98/ 930	Carbamate kinase	Amino acid kinase domain	WP_065949358.1 (100/99)
Orf-1/ <i>orf-1</i>	13903.96/ 363	Acetyl-CoA carboxylase alpha subunit	DUF5064	WP_065895065.1 (100/100)
Orf1/ <i>orf1</i>	54742.5/ 1509	Transcriptional regulator	ACT domain; PAS domain; Sigma-54 interaction domain; PAS domain; HTH domain	WP_065949357.1 (100/100)
Orf2/ <i>orf2</i>	13798.36/ 384	Glycine cleavage system H-protein	Biotinyl lipoyl domain	WP_065949356.1 (100/100)
Orf3/ <i>orf3</i>	102206.59/ 2838	Glycine dehydrogenase (Glycine cleavage system P-protein)	PLP-dependent aspartate aminotransferase (fold-type I) domain	WP_065949355.1 (100/99)
Orf4/ <i>orf4</i>	48961.74/ 1377	L-serine ammonia lyase	L-serine dehydratase alpha and beta chains	WP_065949354.1 (100/99)
Orf5/ <i>orf5</i>	40424.23/ 1125	Aminomethyltransferase (Glycine cleavage system T-protein)	Folate-binding domain; glycine cleavage T-protein C-terminal barrel domain	WP_065949353.1 (100/99)
Orf6/ <i>orf6</i>	7678.65/ 213	Cold-shock protein	DNA-binding domain	WP_065949352.1 (100/100)
Orf7/ <i>orf7</i>	18377.98/ 489	Transmembrane protein	RDD domain	WP_065949351.1 (100/100)
Orf8/ <i>orf8</i>	38252.36/ 1059	Quinolinate synthetase	Quinolinate synthetase A protein	WP_065949350.1 (100/100)

ObaA/ <i>obaA</i>	26341.56/ 714	Transcriptional regulator	Autoinducer-binding domain; DNA-binding HTH domain	WP_065949349.1 (100/99)
ObaB/ <i>obaB</i>	18690.35/ 510	<i>N</i> -acylhomoserine lactone synthase	<i>N</i> -acyltransferase domain	WP_065949348.1 (100/100)
ObaC/ <i>obaC</i>	36033.87/ 951	<i>N</i> -oxygenase	Ferritin-like domain	WP_065936855.1 (100/99)
ObaD/ <i>obaD</i>	22678.28/ 600	4-amino-4-deoxychorismate synthase component II	GATase-1 domain	CRM14850.1 (100/100)
ObaE/ <i>obaE</i>	54701.47/ 1461	4-amino-4-deoxychorismate synthase component I	Anthranilate synthase component I, N-terminal region; chorismate binding domain	CRM14833.1 (100/100)
ObaF/ <i>obaF</i>	46268.45 /1263	Bifunctional 4-aminochorismate mutase/4-aminoprephenate dehydrogenase	Rossmann-fold NADB domain; chorismate mutase type II	WP_065949346.1 (100/100)
ObaG/ <i>obaG</i>	48592.95/ 1323	Putative SHMT	PLP-dependent aspartate aminotransferase (fold-type I) domain	WP_065949345.1 (100/100)
ObaH/ <i>obaH</i>	62078.40/ 1704	4-nitrophenylpyruvate decarboxylase	ThDP-dependent enzyme pyrimidine-binding domain; ThDP-dependent enzyme central domain (2-fold Rossmann fold); ThDP-binding domain	WP_065949344.1 (100/100)
ObaI/ <i>obaI</i>	209339.56/ 5733	Dimodular nonribosomal peptide synthase	Condensation domain (C ₂); AMP-binding domain (A ₂); PP-binding domain (PCP ₂); Thioesterase domain (TE ₂); MbtH-like domain; AMP-binding domain (A ₁)	WP_065949343.1 (100/99)
ObaJ/ <i>obaJ</i>	23689.61/ 645	Isochorismatase	Cysteine-hydrolase domain	WP_065949342.1 (100/100)
ObaK/ <i>obaK</i>	10073.63/ 267	Aryl carrier protein	PP-binding domain (ArCP1)	WP_065949341.1 (100/100)
ObaL/ <i>obaL</i>	26832.77/ 771	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	Rossmann-fold NADB domain	WP_065949340.1 (100/99)
ObaM/ <i>obaM</i>	39744.51/ 1074	DAHP synthase	DAHP synthetase class II domain	WP_065949339.1 (100/100)
ObaN/ <i>obaN</i>	52693.61/ 1419	Isochorismate synthase	Chorismate-binding domain	WP_065949338.1 (100/99)
Orf9/ <i>orf9</i>	72153.72/ 1914	Threonine tRNA ligase	TGS domain; tRNA SAD domain; HMG-box DNA-binding domain; threonyl-tRNA synthetase class II core catalytic domain; threonyl anticodon-binding domain	WP_065949337.1 (100/100)
Orf10/ <i>orf10</i>	17007.34/ 471	PK cyclase	PYR1-like SRPBCC domain	WP_065949336.1 (100/99)

Orf11/ <i>orf11</i>	NA/73	tRNA-lysine (TTT)	N/A	N/A
Orf12/ <i>orf12</i>	24252.64/ 648	7-carboxy-7-deazaguanine synthase	4e-4s single cluster domain; radical SAM superfamily domain	WP_065949335.1 (100/99)
Orf13/ <i>orf13</i>	21974.62/ 630	Tol/Pal system protein	N-terminal trimerisation domain; BamD superfamily domain	WP_065949334.1 (100/100)
Orf14/ <i>orf14</i>	45916.31/ 1260	Tol/Pal system translocation protein	TolB amino-terminal domain; WD40-like beta propeller repeat (x3)	WP_065887603.1 (100/100)
Orf14/ <i>orf14</i>	38686.91/ 1074	TolA-like protein	TonB family C-terminal domain	WP_034115767.1 (100/100)

Table 3.3. PCSs identified in, and flanking, the putative obafluorin (1) BGC.

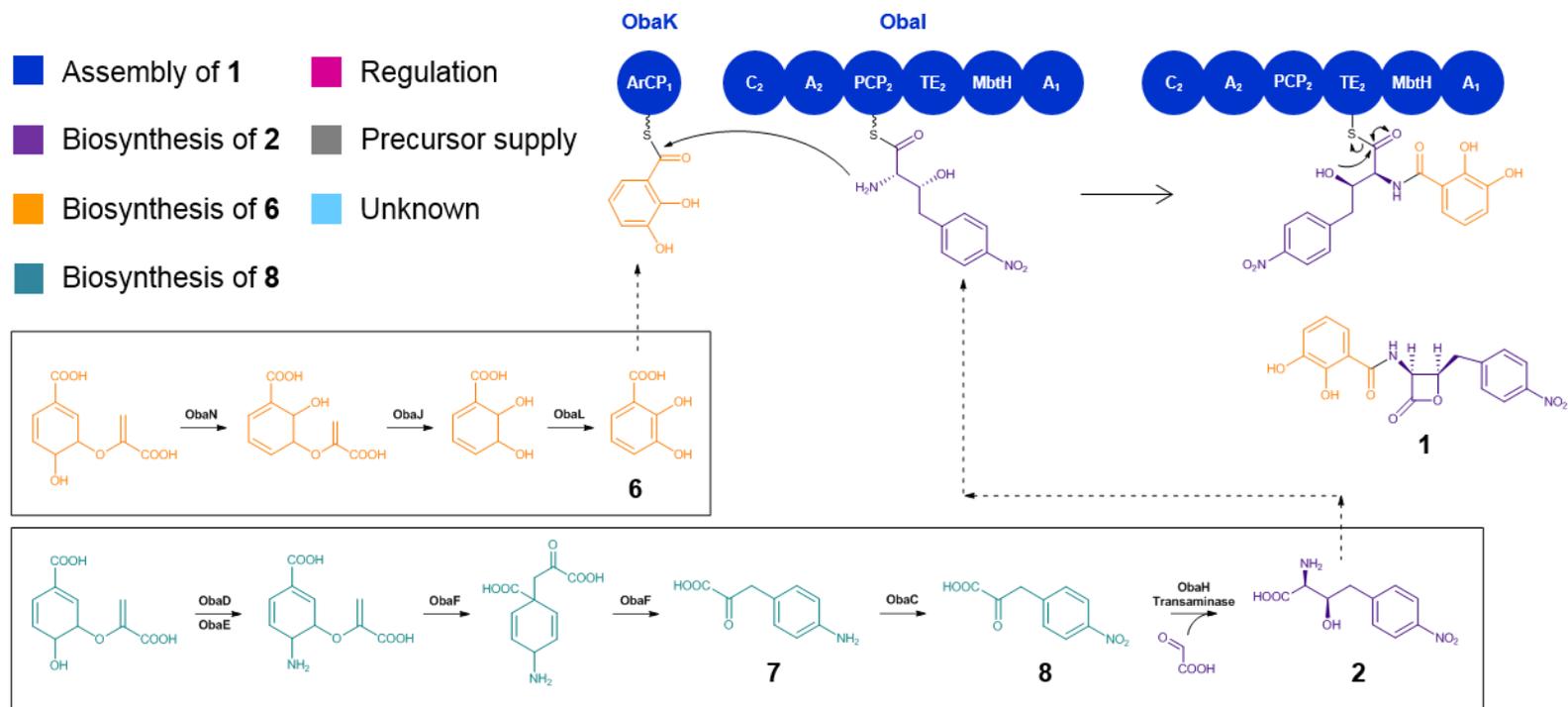
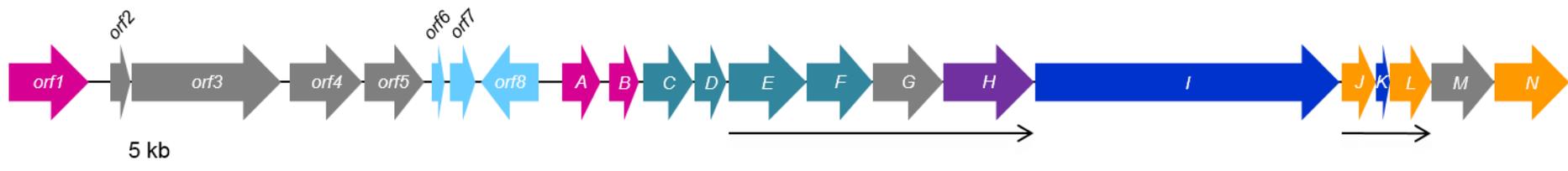
Genes are annotated in accordance with Figure 3.5.

3.5.1 Biosynthesis of 2,3-DHBA (6)

Bioinformatic analysis of the *oba* BGC, suggested that three PCSs encode for the well understood pathway to **6** from chorismate (Figure 3.6a), a major branch point in the shikimic acid pathway to proteinogenic aromatic amino acids (Knaggs, 2003). *obaN* encodes an isochorismate synthase, which would act first to convert chorismate to isochorismate. The subsequent activities of *obaJ*, an isochorismatase, and *obaL*, a 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, should then be sufficient to produce **6** (Figures 3.5 and 3.6a). **6** is a common constituent of bacterial siderophores (Hider and Kong, 2009) due to the ability of its deprotonated catechol group to coordinate metal ions. Homologues of *obaJLN* have previously been identified in the biosynthesis of enterobactin (Walsh *et al.*, 1990), bacillibactin (May *et al.*, 2001) and vibriobactin (Keating *et al.*, 2000), which all comprise three catechol moieties for the coordination of ferric iron (Figure 3.6b).

3.5.2 Biosynthesis of 4-APP (7)

obaDEF also encode previously characterised functions in the chloramphenicol (He *et al.*, 2001) and pristinamycin (Blanc *et al.*, 1997) pathways, and should be necessary for the production of **7**, another chorismate-derived intermediate (Figures 3.5 and 3.7). **7** (or **8** derived from **7**) is the proposed substrate of a ThDP-dependent enzyme during the biosynthesis of **2** (Figure 3.1). *obaD* and *E* are homologues of *pabA/trpG* (49% and 45% shared amino acid identity respectively) and *pabB/trpE* (31% and 28%) (Slock *et al.*, 1990), which in primary metabolism, encode the two components of a 4-amino-4-deoxychorismate (ADC) synthase. This enzyme converts chorismate to ADC, and associates with PabC (Nichols *et al.*, 1989), an ADC lyase,



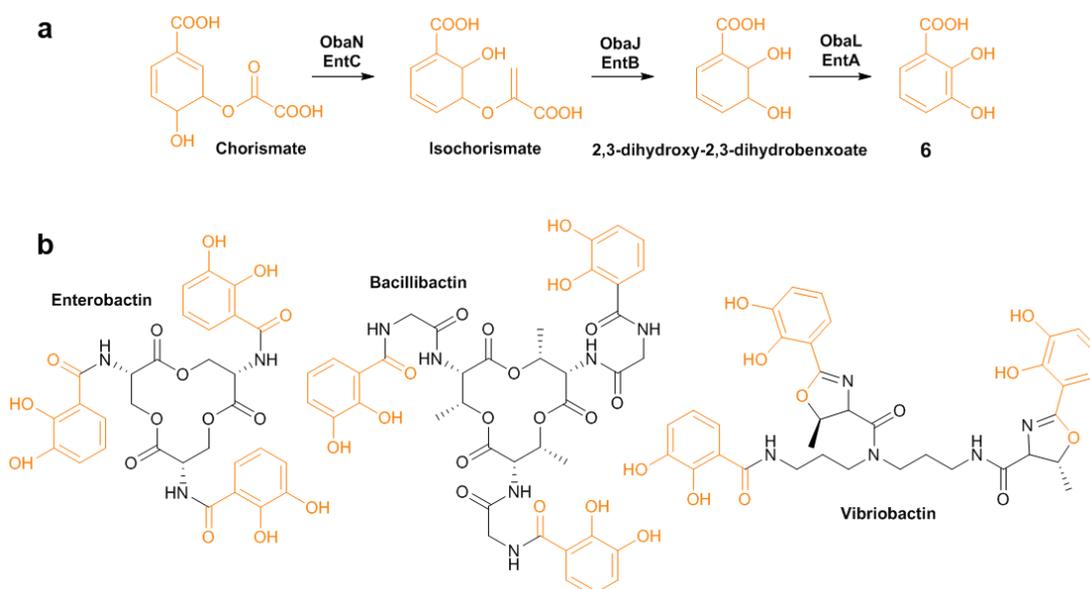


Figure 3.6. Proposed biosynthesis of 2,3-DHBA (6) and catechol-comprising NPs. (a) The proposed biosynthetic pathway to **6** catalysed by the products of *obaJLN*. Homologues in the enterobactin biosynthetic pathway are also depicted. (b) Examples of NP siderophores that incorporate **6** during their biosynthesis. **6** moieties are highlighted in orange, consistent with Figure 3.5.

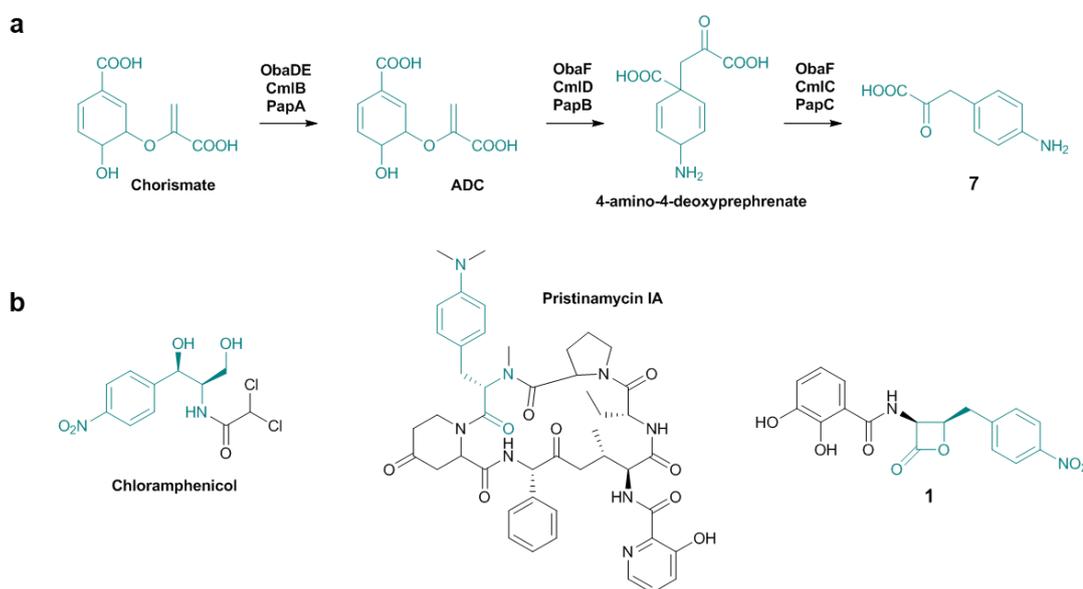


Figure 3.7. Proposed biosynthesis of 4-APP (7) and NPs with 4-APP-derived moieties. (a) The proposed biosynthetic pathway to **7** catalysed by the products of *obaDEF*. Homologues in the chloramphenicol (*Cml*) and pristinamycin IA (*Pap*) biosynthetic pathways are also depicted. (b) Examples of NPs that incorporate **7**-derived moieties - highlighted in turquoise, consistent with Figure 3.5.

to generate 4-aminobenzoic acid (4-ABA) for incorporation into folic acids. In chloramphenicol and pristinamycin biosynthesis, the two ADC synthase components are fused as a single enzyme encoded by *cmIB* and *papA* respectively (Figure 3.7a). In these specialised metabolic contexts ADC is not converted to 4-ABA, but is instead converted to 4-amino-4-deoxyprephenate by an ADC mutase, before final conversion to **7** by a 4-aminoprephenate dehydrogenase. These functions are encoded in the bifunctional product of *obaF* (Figure 3.5 and 3.7a), and in the discrete genes *cmID/papB* and *cmIC/papC* in the chloramphenicol and pristinamycin BGCs.

3.5.3 Biosynthesis of AHNB (**2**)

In accordance with Herbert and Knaggs' hypothesis for the biosynthesis of **2** (Herbert and Knaggs, 1990), a ThDP-dependent enzyme is encoded in the proposed *oba* BGC by *obaH*, which is annotated as a phenylpyruvate decarboxylase (pPDC). This type of activity was proposed to be responsible for catalysing C-C bond formation between **7** (or **8**) and glyoxylate to generate an acyloin intermediate of **2** (Figure 3.1). The *obaH* encoded amino acid sequence was submitted for analysis using the Phyre2 server (Kelley *et al.*, 2015), and the highest structural match was to a pyruvate oxidase from *E. coli* with 95% coverage and 100% confidence. Intriguingly, there is no transaminase for acyloin transamination encoded in close proximity to the *oba* BGC cluster (Figure 3.5), which would imply that a background enzyme is able to fulfil the function of introducing the α -amino group of **2**, if the previous proposal was correct (Herbert and Knaggs, 1990). In both chloramphenicol and pristinamycin biosynthesis (Blanc *et al.*, 1997; He *et al.*, 2001), the transaminase required for introducing the aminophenyl group of **3** is also not encoded in either BGC.

3.5.4 Obafluorin (**1**) dipeptide assembly

At the centre of the *oba* BGC is the dimodular NRPS-encoding gene *obal* (Figure 3.5), the product of which was predicted to catalyse formation of the amide bond between **2** and **6**. The thioesterase (TE) domain of Obal is unusually located between the peptide carrier protein (PCP₂) domain and first adenylation (A₁) domain, which could catalyse release of the resulting enzyme bound *pseudo*-dipeptide with concomitant formation of the β -lactone moiety to yield **1**. This was proposed because no *oleC* homologue (Christenson *et al.*, 2017) is present in the **1** BGC. Analysis of the A-domain sequences of Obal using NRPSpredictor2 (Röttig *et al.*, 2011) identified **6** as the likely substrate of A₁, whereas L-threonine was predicted to be the substrate of A₂.

The domain architecture of Obal is further unusual as the putative A₁-domain is located at the C-terminus of the enzyme, rather than towards the N-terminus as would be expected. Moreover, A₁ is adjacent to an embedded MbtH-like protein domain. MbtH-like proteins are auxiliary proteins required for the activity of many NRPS A domains (Quadri *et al.*, 1998a; Felnagel *et al.*, 2010; Baltz, 2011). Some A domains require MbtH-like proteins for their functional expression, suggesting MbtH-like proteins may have a chaperone-like function (Heemstra *et al.*, 2009). Embedded MbtH-like domains like that observed in the product of Obal are rare, and discrete MbtH-like proteins are more commonly observed. Embedded MbtH-like domains in NRPSs have been reported for only two other BGCs: those responsible for the biosynthesis of nikkomycin (Chen *et al.*, 2002) and streptolydigin (Herbst *et al.*, 2013). During streptolydigin biosynthesis, an A-MbtH-like didomain protein is involved in the activation of 3-methylaspartate. Whilst the amino acid residues critical for inter-domain interaction between these fused domains are well characterised at the structural level, no biochemical data exists to verify that *in cis* domains are functionally essential *in vivo*. The exact role of MbtH-like proteins in NRPS catalysis is also yet to be determined.

Obal also contains only a single PCP₂ domain where two such domains would be expected for its correct function. Further analysis of the BGC suggested that *obaK* encodes the 'missing' discrete aryl carrier protein (ArCP₁). The presence of an ArCP is anticipated as this protein function would be required for the tethering of **6** (following activation by the A₁-domain) to participate in amide bond formation. Despite, their unusual organisation, all of the domains necessary for the assembly of **1** are accounted for by ObalK (Figure 3.5).

3.5.5 Nitro group formation

The gene *obaC* encodes a putative non-heme di-iron monooxygenase related to the arylamine oxygenases AurF from the aureothin pathway (73% query cover, 25% identity) (He and Hertweck, 2001; Choi *et al.*, 2008) and CmlL from the chloramphenicol pathway (72% query cover, 19% identity) (Lu *et al.*, 2012; Knoot *et al.*, 2016). This enzyme was predicted to install the nitro group of **1**. Although the timing of *N*-oxygenation during the biosynthesis of **1** cannot be determined based on *in silico* data alone, both AurF and CmlL catalyse the conversion of an aromatic amine to a nitro group at intermediate stages in their respective biosynthetic pathways (Figure 3.8a and b), so ObaC activity is illustrated at an earlier stage of **1** biosynthesis in Figure 3.5. However, ObaC could use one of several pathway intermediates as its

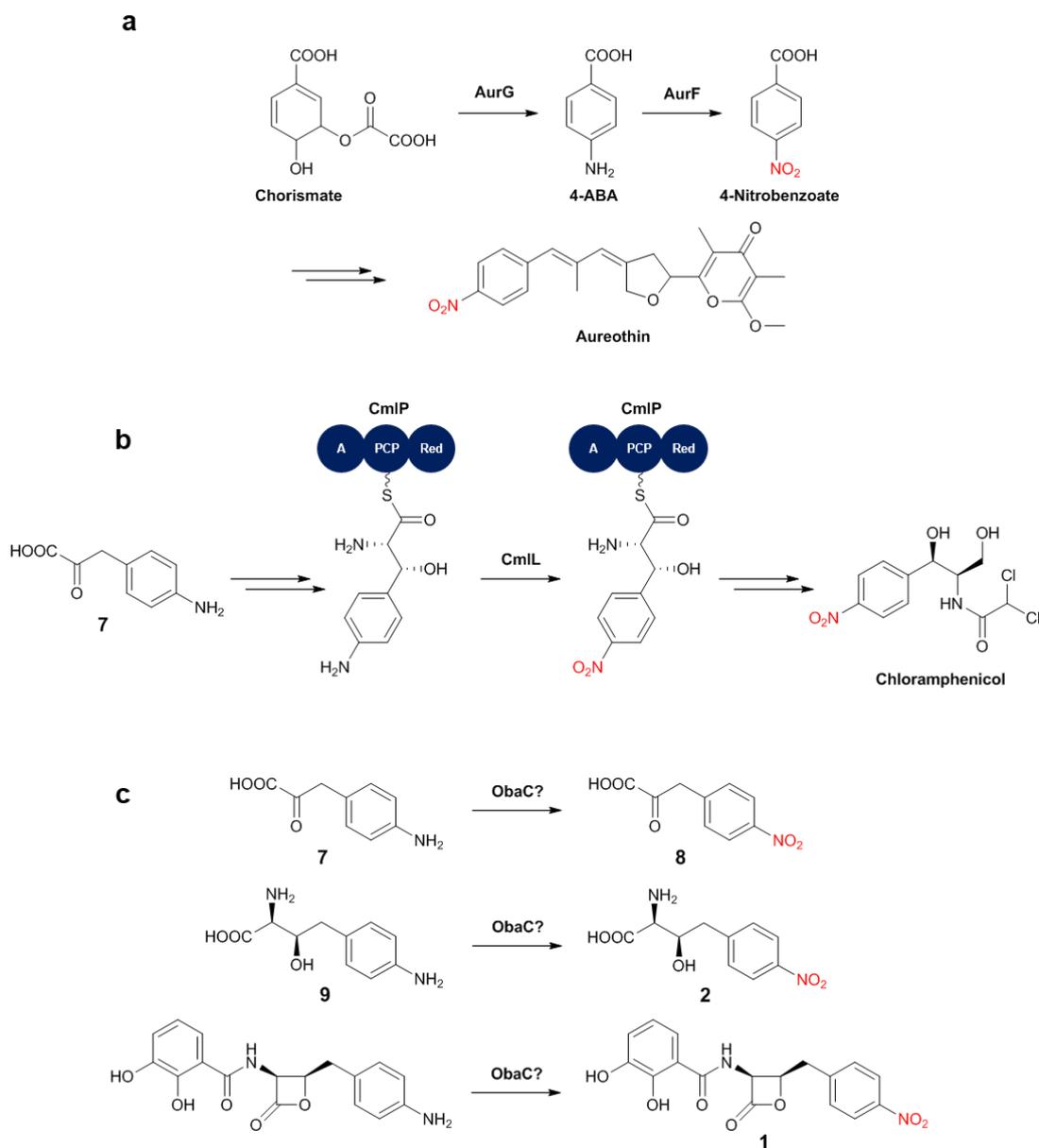


Figure 3.8. N-oxygenation during NP biosynthesis. (a) N-oxygenation of 4-ABA catalysed by AurF during aureothin biosynthesis. (b) N-oxygenation of 4-amino- β -hydroxyphenylalanyl-S-PCP by CmlL during chloramphenicol biosynthesis. (c) Potential substrates of the N-oxygenase ObaC during the biosynthesis of 1.

specific substrate, as illustrated in Figure 3.8c, and experimental work will be required to confirm which it the true substrate.

3.5.6 Precursor supply

With all the biosynthetic functions necessary for **1** production seemingly accounted for, several unassigned PCSs in the *oba* BGC remained that could potentially be involved in supplying important precursors to biosynthetic enzymes in the **1** pathway (Figure 3.5). *obaM* for example, encodes a putative type II DAHP synthase, the first enzyme in the shikimic acid pathway (Figure 3.9). Homologues of this gene have been observed in a number of bacterial BGCs (Blanc *et al.*, 1997; He *et al.*, 2001; Gulder and Moore, 2010), where their products are believed to mitigate the aromatic amino acid-based negative feedback of primary metabolic type I DAHP synthases (Knaggs, 2003), and drive flux into specialised metabolism. This indicates a likely role for ObaM in chorismate supply for the biosynthesis of both **2** and **6**.

ObaG encodes a putative SHMT, a ubiquitous PLP-dependent enzyme that catalyses the interconversion of L-serine and glycine (Figure 1.11), using tetrahydrofolate (THF) as a C₁ group acceptor (Schirch *et al.*, 1985). Given that Herbert and Knaggs demonstrated that glycine (via glyoxylate) is the likely source of the C-1 and C-2 in **2** (1990; 1992b), it seemed possible that this additional copy of SHMT was involved in driving flux through the ObaH-catalysed reaction by supplying an excess of glyoxylate. ObaM would work in parallel, supplying chorismate for the biosynthesis of **6** and **7**, the other substrate of the proposed ObaH-catalysed reaction. The presence of an intact glycine cleavage system (GCS) encoded by *orf2345* upstream of the main *oba* cluster made a role for ObaG in precursor supply more convincing, because it too can operate reversibly to generate glycine from CO₂, NH₄⁺ and 5,10-methylene-THF (Kikuci *et al.*, 2008), potentially operating in tandem with ObaG to supply ObaH with glycine (glyoxylate) (Figure 3.10).

3.5.7 Regulation of obafluorin (1) biosynthesis

obaA and *obaB* encode LuxR and LuxI homologues respectively, indicating that the biosynthesis of **1** is regulated in a quorum sensing (QS)-dependent manner. Bacteria are able to synchronise gene expression and particular behaviours as a function of population density via the release of small signalling molecules called autoinducers (Camilli and Bassler, 2006). These alter gene expression when they are detected above a minimum threshold stimulatory concentration. In Gram-negative bacteria, the predominant signalling molecules are typically *N*-acylhomoserine lactones (AHLs),

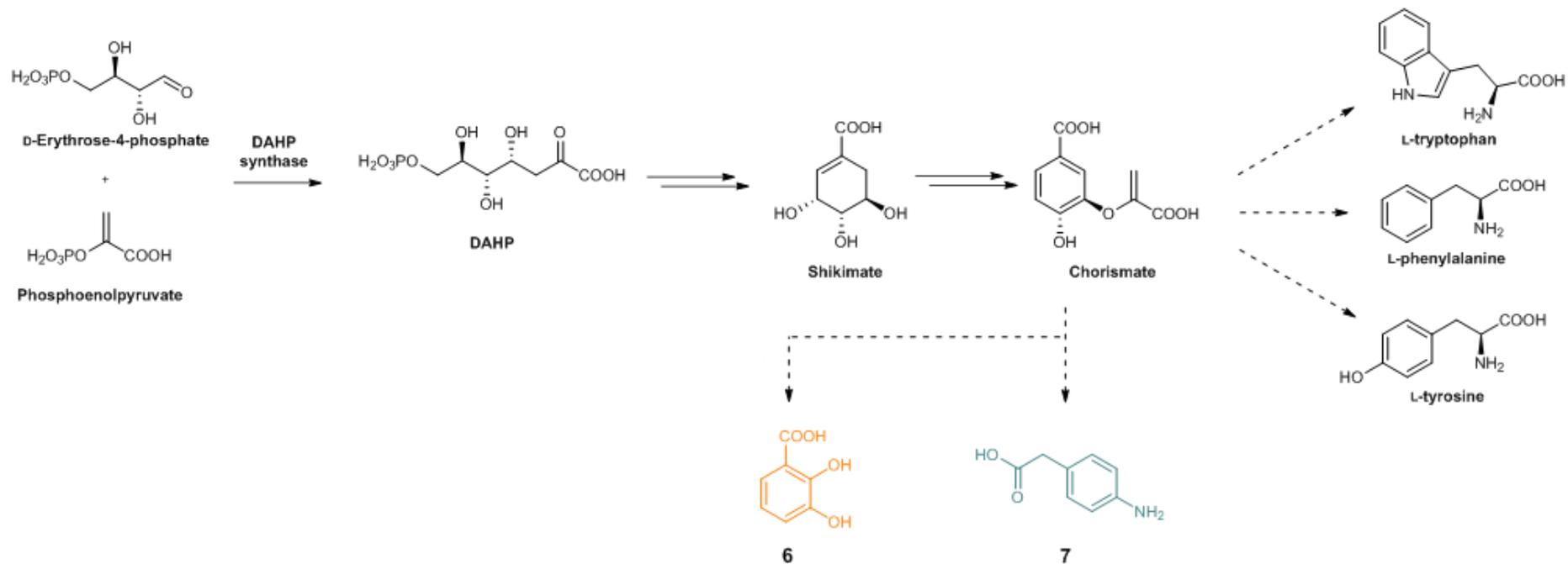


Figure 3.9. Role of DAHP synthase in the shikimic acid pathway to aromatic amino acids. DAHP synthases catalyse the stereospecific condensation of phosphoenolpyruvate and D-erythrose-4-phosphate to form DAHP and are responsible for the amount of carbon entering the shikimic acid pathway.

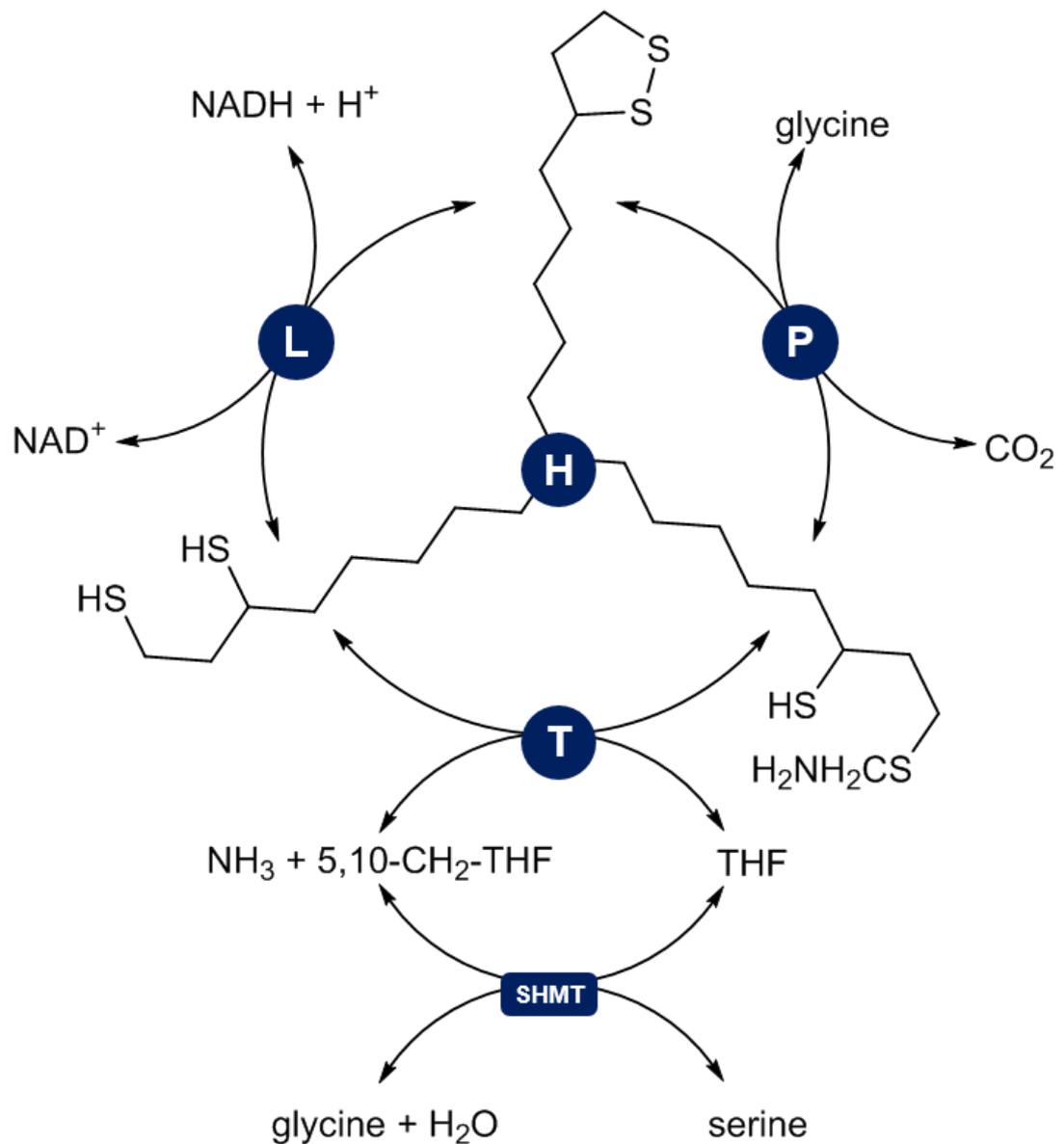


Figure 3.10. Illustration of the reactions catalysed by the glycine cleavage system (GCS). The glycine cleavage system catalyses the following reversible reaction: $\text{glycine} + \text{THF} + \text{NAD}^+ \rightleftharpoons 5,10\text{-methylene-THF} + \text{CO}_2 + \text{NH}_3 + \text{NADH}$. The H-protein, modified with lipoic acid, functions to shuttle intermediates between the P-protein (a glycine dehydrogenase), the T-protein, (and aminomethyltransferase) and the L-protein (a dihydrolipoyl dehydrogenase). It is an important part of the glycine and serine catabolic pathways and an important source of C₁ units for THF-dependent metabolism.

and are the product of autoinducer synthases of the LuxI family, named after the first characterised member from the bioluminescent bacterium *Vibrio fischeri*. AHLs all share a common homoserine lactone moiety, but comprise distinct acyl side chains. These side chains are recognised by specific LuxR-type receptor enzymes, the acyl-binding pockets of which exclusively accommodate the acyl-chain of their cognate AHL signal. LuxR-AHL complexes bind DNA promoter regions and activate the transcription of QS-controlled genes. QS is employed by bacteria to regulate processes that require the combined contribution of numbers of cells to be productive, including bioluminescence, biofilm formation, antibiotic production and virulence factor expression as examples (Camilli and Bassler, 2006). Whilst not within the scope of this project, experiments to investigate the regulation of the biosynthesis of **1** are discussed in Chapter 9.

3.5.8 Unassigned PCSs

Several PCSs within the 30 kbp putative *oba* BGC could not be assigned direct roles in **1** production. *Orf6* encodes a cold shock protein, thought to counteract the harmful effects of rapid drops in temperature. They are proposed to act as nucleic acid chaperones, preventing secondary structure formation in mRNA and maintaining normal levels of translation initiation (Jiang *et al.*, 1997). *Orf7* encodes a predicted transmembrane protein belonging to the RDD family, so-called because they contain one highly conserved arginine and two highly conserved aspartate residues. *Orf7* could possibly be involved in **1** export but the roles of RDD family proteins have not been well characterised. *Orf8* encodes a quinolinate synthase, an iron-sulphur enzyme responsible for the biosynthesis of quinolinate, the precursor of the ubiquitous cofactor nicotinamide adenine dinucleotide (NAD) (Ollagnier-de Choudens *et al.*, 2005). NAD is involved in numerous redox reactions through hydride transfer, and the products of both *obaF* and *obaL* comprise a Rossmann fold that could bind it as a cofactor for catalysis. However, a role in cofactor supply is unlikely given there are no enzymes involved in ThDP or PLP biosynthesis associated with the cluster and cofactor biosynthetic genes are not typically found in NP BGCs. Finally, *orf1* encodes a σ -54 transcriptional regulator homologous to the multidomain TyrR protein, a regulator of aromatic amino acid biosynthesis and transport (Pittard *et al.*, 2005). The presence of a DAHP synthase gene *obaM* in the cluster indicates that chorismate supply from the shikimic acid biosynthetic pathway is tightly coordinated in the biosynthesis of **1**, and whilst its presence could be purely coincidental, it is not unreasonable to wonder whether a *tyrR*-like gene might also contribute to the regulation of chorismate supply.

3.5.9 Identifying *oba* BGC boundaries

Upstream of *orf1* are PCSs encoding products involved in arginine and ornithine metabolism including a carbamate kinase, ornithine transcarbamylase, arginine deaminase and two arginine/ornithine antiporters. Downstream of *obaN* are two PCSs followed by a lysine tRNA gene, which has been identified as a common site of integration for genomic islands in *Pseudomonas* spp. (Williams, 2002; Mavrodi *et al.*, 2009), and could well indicate the true downstream boundary of the *oba* cluster. The two PCSs encode a threonyl-tRNA synthetase (ThrRS), responsible for ligating L-threonine to the nucleotide triplet of its cognate tRNA (Sankaranarayanan *et al.*, 1999), and a PK cyclase, enzymes typically responsible for the aromatisation of PKs. Intriguingly, the PK cyclase PCS product shares 59% amino acid sequence identity with MgoD, the missing PCS from the incomplete mangotoxin BGC encoded elsewhere in the *P. fluorescens* ATCC 39502 genome. It is unlikely that these PCSs play any role in the production of **1**, are more likely present by chance, falling within the lysine tRNA recombination site boundary.

3.6 Discussion

High molecular weight genomic DNA was prepared and submitted for sequencing using the PacBio SMRT platform. A single contig representing a single circular chromosome for the strain was obtained and was found to closely resemble the architecture and composition of other representative *P. fluorescens* strains. ATCC 39502 was identified as a more 'talented' producer of NPs with 17 BGCs identified, compared to around ten in other strains. NRP BGCs were well represented with eight being identified and amongst those the likely *oba* cluster was readily identified (Figure 3.5).

Whilst several of the putative BGCs identified in the ATCC 39502 genome were assigned to known NPs such as pseudomonine and pyoverdine, the majority are orphan clusters for which the product has not been identified. Predictions were also possible for two poeamide-like BGCs. Perhaps the most unusual putative BGC was a cluster comprising a dimodular class I terpene synthase/cytochrome P450 enzyme, a class I SAM methyltransferase and an IPP isomerase. While around 50,000 terpenoid metabolites representing nearly 400 distinct families have been isolated from plants and fungi, only a minor fraction have been identified in prokaryotes (Yamada *et al.*, 2015). Their discovery and characterisation is challenging because bacterial terpene synthases share low amino acid sequence similarity to fungal and plant enzymes, and typically show low levels of mutual sequence similarity. Neither

terpenoid NPs nor terpene synthases have been described in the pseudomonads, but these three clustered genes appear to be highly conserved among a number of *Pseudomonas* spp. and perhaps warrant further investigation. Blastp analysis of the terpene synthase domain identified that all the conserved residues for substrate binding and for chelation of 3Mg^{2+} ions, responsible for ionisation of DMAPP (Oldfield and Lin, 2012), are present. In order to determine whether this cluster (minimally 3.6 kbp) is functional and does produce a terpenoid NP, it could be cloned for expression in a heterologous host such as *P. putida*, which is known for its metabolic versatility and which does not contain the cluster (Loeschcke and Thies, 2015).

The proposed *oba* BGC comprises PCSs encoding products for the biosynthesis of **6** (*obaJLM*) and **7** (*obaDEF*) from chorismate. These have previously been well characterised in the biosynthetic pathways to several siderophore NPs, and to chloramphenicol and pristinamycin I and II respectively. Central to this project is the presence of a ThDP-dependent pPDC, *obaH*, proposed by Herbert and Knaggs to catalyse an C-C bond forming reaction between glyoxylate and **7** to generate **2** (Figure 3.1).

obaI and *obaK*, encode an NRPS and discrete ArCP respectively, which between them comprise all the domains deemed necessary to assemble **1** from **2** and **6**, and form the β -lactone ring of the final product. The unusual organisation and MbtH-like domain in the product of *obaI* is intriguing. On the basis of these analyses one might hypothesise that the noncanonical positioning of A-domains within the assembly line provides them both access to the embedded MbtH-like domain, a potential prerequisite for efficient catalysis. This organisation likely precludes functional interaction of the A_1 domain and the condensation (C_2) domain, a consequence of which is the requirement for a discrete, and therefore mobile, ArCP₁ (*ObaK*). This would enable both interaction with the A_1 domain for acylation with **6**, and subsequent interaction with the C_2 -domain for peptide bond formation. It would be important to query this possibility as *obaK* appears to have evolved by the splitting of an ancestral *entB*-like gene, which usually encode for didomain proteins involved in the biosynthesis of **6** (Walsh *et al.*, 1990; Drake *et al.*, 2006). Indeed, the adjacent *obaJ* gene corresponds to the N-terminal isochorismatase domain of EntB (Gehring *et al.*, 1997) and is required for production of **6**.

The final gene that could be assigned a biosynthetic function was *obaC*, which encodes an *N*-oxygenase likely to be responsible for oxygenation of the **1**

phenylamino group to a nitro group. However, on what intermediate this *N*-oxygenation occurs is unclear. The cluster also encodes several products which could play roles in driving flux through the biosynthetic pathway via the supply of important precursor molecules – chorismate in the case of *obaM* and glyoxylate from glycine in the case of *obaG* and *orf2345*. Finally, the biosynthesis of **1** appears to be regulated in a QS-dependent manner, due to the presence of *luxI* and *R* homologues *obaA* and *B*, but no experimental work was performed in this project to explore the regulation.

Whilst the *in silico* data described above allow for a hypothesis for the biosynthesis of **1** to be made, experimental validation was required to confirm the involvement of PCSs described and to provide further information about the temporal organisation of biosynthesis (e.g. timing of *N*-oxygenation). Critically, it would allow the role of the *obaH* product in the pathway to be characterised and the mechanism proposed by Herbert and Knaggs (1990) for the biosynthesis of **2** to be validated.

Chapter 4:
Mutational analysis of the *oba*
BGC

Chapter 4: Mutational analysis of the *oba* BGC

4.1 Introduction

In order to characterise NP biosynthesis at the molecular level, a range of different approaches can be used. The majority of methods employ comparative metabolic profiling to detect the presence or absence of NPs and their intermediates under different experimental conditions, often following the disruption of biosynthetic loci (Zerikly and Challis, 2008). For genetically intractable organisms, the cloning and expression of entire BGCs in heterologous hosts can link an NP to its BGC, through the detection of non-native metabolites in the host background (Zhang *et al.*, 2016). To this end, 'super-hosts' have been developed in which highly expressed native NP BGCs are removed to aid the identification compounds biosynthesised from introduced BGCs (Gomez-Escribano and Bibb, 2011). Removal of native BGCs has the added benefit of increasing the availability of primary metabolites for biosynthesis in the heterologous host. A similar result has recently been achieved by deleting *hfq* in *Photorhabdus* and *Xenorhabdus* spp., which causes the majority of native NP BGC to become post-transcriptionally 'silent' (Tobias *et al.*, 2017). *In vitro* reconstitution of biosynthetic pathways mitigates the need for a host all together and removes the restrictions of regulatory mechanisms. However, problems can be encountered in achieving the expression of multiple soluble biosynthetic proteins, and in obtaining complex precursors required as substrates.

Genetic tools for *P. fluorescens* are reasonably well established, allowing for the comparison of WT and Δoba strains to identify genes responsible for production based on the presence or absence of particular metabolites. Given the assembly-line paradigm of NP biosynthesis, in some instances it is possible to detect the immediate substrate of a disrupted biosynthetic enzyme, which accumulates because it is no longer utilised. In this way, a top-down approach can be used to dissect biosynthetic pathways based on the accumulation of pathway intermediates. Many intermediates are often unstable or recycled by primary metabolic pathways and so are not detectable, which can limit this approach.

This chapter describes the use of mutagenesis to dissect the biosynthetic pathway for obafluorin (**1**). First, suitable **1** production conditions and analytical methods to define a measurable 'phenotype' were established and pure **1** was obtained for use as a reference in experiments. Suicide and complementation vectors were

constructed/modified to facilitate gene disruption and complementation of *oba* PCSs, and a series of mutagenetic experiments were then performed to disrupt different elements of the biosynthesis of **1**. Mutagenesis of genes suspected to be involved in precursor supply was also performed to evaluate their necessity for production. Where possible, mutants were also complemented chemically to identify likely pathway products. Based on the accumulation of shunt metabolites and intermediates, predictions are made about the function and timing of different biosynthetic enzymes during the biosynthesis of **1**.

4.2 Analysis of obafluorin (**1**) production

P. fluorescens was grown according to reported production conditions (Wells *et al.*, 1984), though soil extract was omitted from the OPM recipe and dH₂O was used instead. LCMS/MS allowed the identification of the corresponding UV peak for **1**. Cultures were then scaled up and 13 mg of **1** were purified (Materials and Methods section 2.10.1). Dr. Zhiwei Qin, a postdoctoral researcher in the group, subsequently subjected the purified sample to HPLC, HR-LCMS and NMR analyses, and all data matched published values (Tymiak *et al.*, 1985). An analytical reverse-phase HPLC method was then established to allow facile detection of **1** production from small-scale 100 mL cultures (Materials and methods section 2.10.2). **1** appears as a dominant peak when chromatograms are generated for absorbance at 270 nm (absorption maxima at 215 nm and 260 nm) with a retention time of ~10 min, along with five minor peaks that elute at slightly lower organic solvent concentrations (Figure 4.1). **10** represents the hydrolysed ring-open form of **1** that is always present in reference samples due to the ease with which the β -lactone ring is hydrolysed, and this was verified by HR-LCMS. Two of the peaks were identified as 4-NPE (**4**) and 4-NPAA (**5**) based on detection of exact masses by HR-LCMS and HPLC comparison with authentic standards. These compounds were also detected by Herbert and Knaggs (1988, 1990). Synthetic standards for AHNB (**2**), 4-APA (**3**), 2,3-DHBA (**6**), 4-NPP (**8**), AHAB (**9**) and 4-ABA were acquired or synthesised, but did not account for peaks **A** and **B**. These peaks could also not be isolated or detected by HR-LCMS. Synthetic standards of both **2** and **9** were prepared by Prof. Barrie Wilkinson and Dr. Daniel Heine in the group (see Materials and Methods 2.10.1). At this point the production phenotype for **1** was sufficiently characterised, with reference material for **1** and a broad range of plausible pathway intermediates and break down products acquired for comparison with Δ *oba* gene strain profiles.

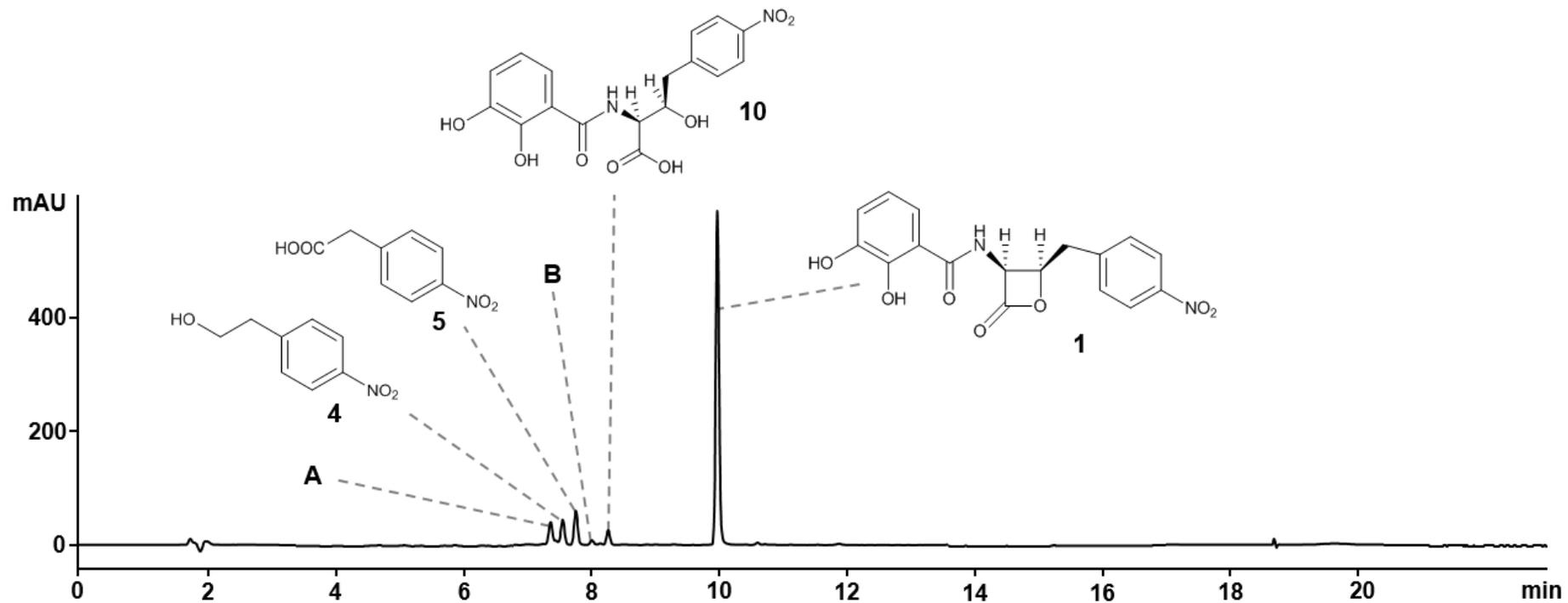


Figure 4.1. HPLC profile for a WT obafuorin (1) production culture extract at 270 nm. Verified peak identities are linked to the molecule they represent. Peaks A and B are as yet uncharacterised.

4.3 Construction of vectors for mutational analysis

4.3.1 pTS1 suicide vector construction

To generate gene disruptions in predicted *oba* PCSs, Dr. Jacob Malone (JIC) kindly provided the ColE1-based suicide vector pME3087 (Voisard *et al.*, 1994), for performing knockouts in *Pseudomonas* spp. This vector works by double homologous recombination using flanking regions of the gene for disruption. The first homologous recombination event that integrates the vector into the host genome is selected for with tetracycline, as a Tc^R cassette is encoded in the pME3087 backbone. Following a PCR-based screen for positive colonies, selected clones are grown through several generations in the absence of antibiotic to allow for a second recombination event to occur to excise the vector backbone, resulting in either a) a disrupted copy of the target gene; b) a revertant to the WT sequence; or c) the single-cross over condition being retained. Growth of cells that have lost the vector backbone is arrested by addition of a bacteriostatic concentration of tetracycline, and those that retain the backbone and continue to replicate due to the possession of the Tc^R cassette are subsequently killed by applying phosphomycin and piperacillin. Colonies that grow when subsequently plated can then be screened for tetracycline sensitivity and loss of the vector backbone. In a side project investigating a potential virulence target in *P. aeruginosa* PA01, the antibiotic enrichment for double recombinant strains was not very efficient and >100 colonies had to be screened in order to identify a *bona fide* deletion colony. A counter-selectable marker was therefore introduced into pME3087 to improve selection efficiency. pME3087 was first sequenced *de novo* by primer walking (submitted the sequence to GenBank under the accession no. KX931446) to facilitate cloning of an appropriate marker gene.

The *Bacillus subtilis* *sacB* gene, which encodes a levansucrase, is a popular marker for counterselection in Gram-negative bacteria, as strains carrying this gene die when grown on media containing sucrose (Gay *et al.*, 1985). The mechanism of toxicity is not entirely understood but the accumulation of levans, high molecular weight fructose polymer products of SacB, in the periplasm of Gram-negative bacteria is thought to be responsible. Whilst a putative levansucrase is encoded in the *P. fluorescens* ATCC 39502 genome, it did not appear to impede growth when the strain was grown on 7% sucrose LB agar. The *sacB* gene was therefore cloned as a 2,009 bp *EcoRI* fragment from the site-specific excision vector pFLP2 (Huoang *et al.*, 1998) into pME3087, and successful clones were confirmed by colony PCR. Only clones possessing the *sacB* fragment oriented in which the direction of transcription was away from the pME3087 multiple cloning site (MCS) were selected to prevent unwanted read-through. An *NdeI*

site introduced in the *EcoRI* fragment containing *sacB* was then removed by partial digestion and religation following treatment with Klenow DNA Polymerase (Materials and methods section 2.4.5), leaving a single *NdeI* site in the MCS.

This vector was further tailored by introduction of an expanded MCS. Sense and antisense oligonucleotides (Appendix 1 – Supplementary Table 1) comprising six additional restriction sites (*AvrII-MfeI-NcoI-BmtI-XhoI-NdeI*) were ligated together and cloned into the pME3087 *HindIII* site, which was regenerated upstream of the new MCS. Intended modifications of the vector were confirmed by Sanger sequencing. *XbaI*, *AvrII* and *BmtI* do not cut within the *oba* BGC and so could be used for cloning *oba* PCS flanking regions for knockouts. The use of three restriction enzymes for all knockout constructs facilitates subcloning between them to create different combinations of flanking sequences. This new pME3087-derived suicide vector was named pTS1 (Figure 4.2a) and its sequence was also submitted to GenBank (accession no. KX931445). Figure 4.3 illustrates the efficiency of pTS1-mediated double crossover for selected *oba* PCS disruptions. Following the second recombination step all colonies tested were sensitive to tetracycline, indicative of successful loss of the vector backbone and the generation of WT revertant or genuine deletion colonies. Design of flanking regions for *oba* locus disruptions using this vector are described in the Materials and methods (section 2.5.1).

4.3.2 pJH10TS complementation vector construction

To ensure that effects of a particular gene disruption were non-polar, Prof. Christopher Thomas (University of Birmingham) generously supplied pJH10 (El-Sayed *et al.*, 2003), a self-replicating, broad range IncQ expression vector with an IPTG-inducible promoter that is constitutively expressed in *P. fluorescens* (Mattheus *et al.*, 2010). This vector, and its derivatives, have been used extensively in his lab to complement mutants in the mupirocin biosynthetic pathway (El-Sayed *et al.*, 2003; Hothersall *et al.*, 2007). The MCS was expanded by ligating sense and antisense oligonucleotides (Appendix 1 - Supplementary Table 1) comprising five additional restriction sites (*MfeI-PacI-BmtI-XhoI-NdeI*) together and cloning them into the pJH10 *EcoRI* site (regenerated upstream of the expanded MCS). The resulting complementation vector was named pJH10TS (Figure 4.2b). All alterations were confirmed by Sanger sequencing.

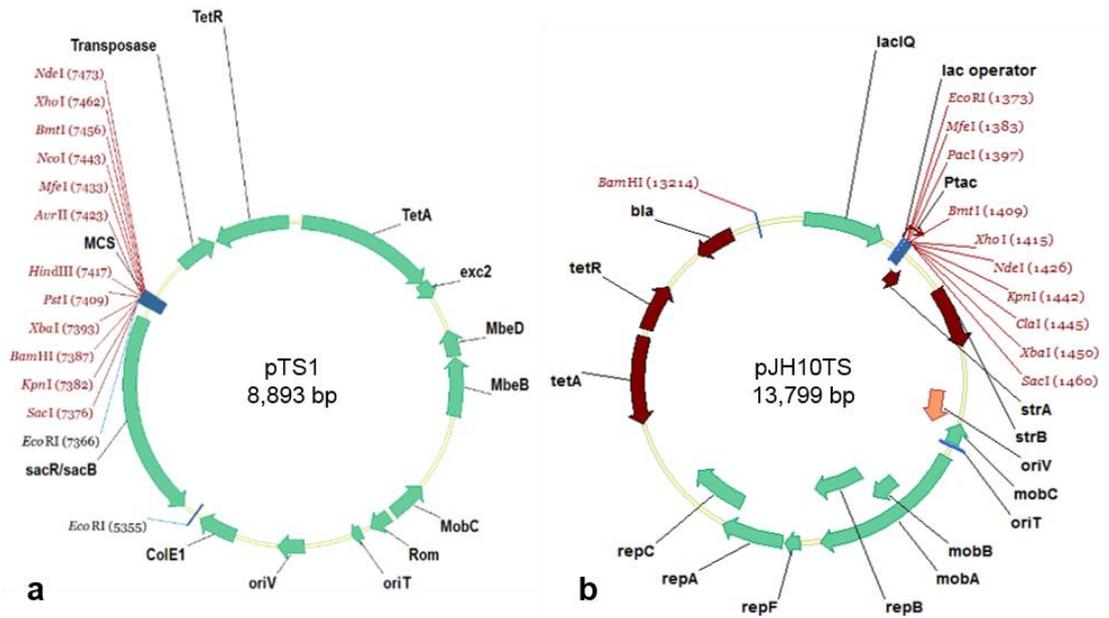


Figure 4.2. Vector maps for the suicide and complementation vectors constructed in this work. (a) pTS1 and (b) pJH10TS. Maps generated using Vector NTI®. Restriction sites in red represent unique sites within the vector.

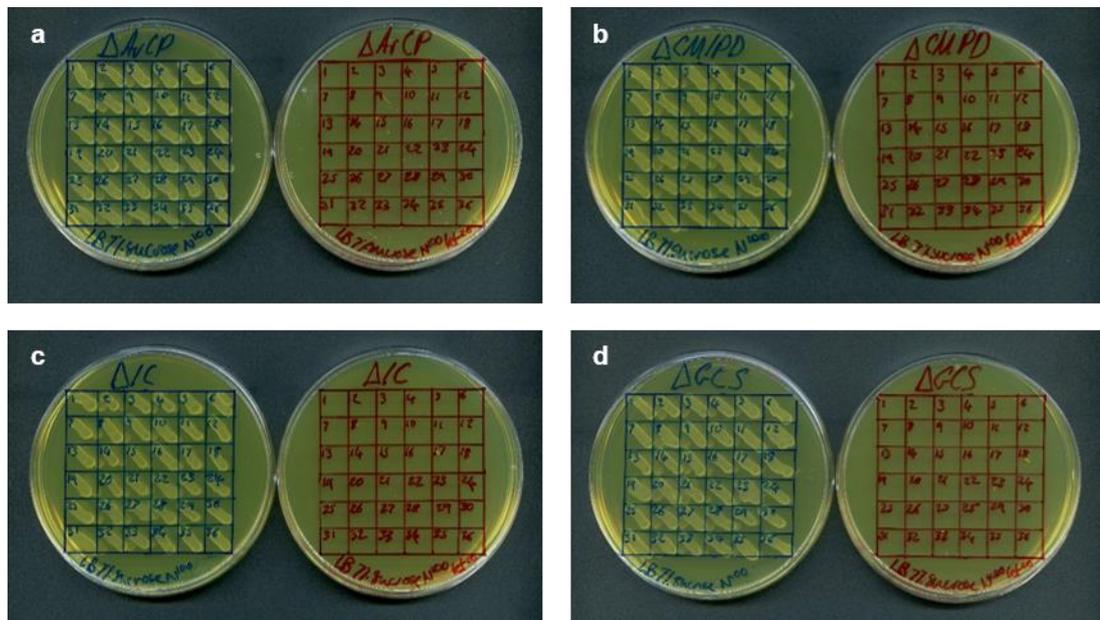


Figure 4.3. Examples of pick and patch plates to test for tetracycline sensitivity following pTS1-mediated gene disruption. Colonies were patched on to LB 7% sucrose N^{100} agar $\pm Tc^{50}$ (- are blue and on the left in each pair, + are red and on the right). (a) Potential Δ obaK mutants (ArCP = aryl carrier protein), (b) Potential Δ obaF mutants (CM/PD = chorismate mutase/prephenate dehydrogenase), (c) Potential Δ obaJ mutants (IC = isochorismatase) and (d) Potential Δ orf2345 mutants (GCS = glycine cleavage system).

4.4 Mutational analysis of *oba* loci

Gene inactivation experiments were performed by generating in frame deletions using pTS1 and all resulting mutants were verified by PCR and Sanger sequencing. At least three independent mutants were checked for each gene deletion. All mutants were functionally confirmed and checked for the lack of polar effects by genetic complementation through ectopic expression of the deleted gene using pJH10TS. As required, chemical complementation with putative pathway intermediates was carried out in parallel by the addition of exogenous material to growing cultures. All strains and their complemented derivatives were grown in triplicate under **1**-producing conditions (alongside the wild type (WT) strain), extracted with ethyl acetate and the extracts subjected to HPLC and HPLC-MS analysis (see Materials and Methods sections 2.5.5 and 2.10.2). Results in this chapter are described using HPLC profile data only, but HR-LCMS data were also acquired for confirmation of identified peaks. A revised biosynthetic pathway based on mutagenesis data presented in subsequent sections is illustrated in Figure 4.4.

4.4.1 Disrupting the biosynthesis of **2,3-DHBA (6)**

Bioinformatic analysis suggested that *obaJLN* encode for proteins involved in the well understood pathway to **6** from chorismate (Figures 3.6 and 4.4). Deletion of either *obaJ* (an isochorismatase) (Figure 4.5) or *obaL* (a 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase) (Figure 4.6) led to the loss of **1** production and four to five-fold elevated levels of **4** and **5**, which are proposed shunt metabolites of the biosynthesis of **2** (Herbert and Knaggs, 1988). The addition of exogenous **6** (0.2 mM final conc.) was sufficient to restore the biosynthesis of **1** to 71% and 68% of WT titres for Δ *obaJ* and Δ *obaL*, respectively. Ectopic expression of the deleted gene in their respective deletion backgrounds also restored **1** production, to 73% (*obaJ*) and 85% (*obaL*) of WT titres. By abolishing production of **6** (and therefore the pathway endpoint **1**), it was anticipated that the predicted pathway intermediate **2** would accumulate (Figure 4.4). However, after exhaustive LCMS analysis and comparison with synthetic references, accumulation of **2** was not detected, nor was the potentially related pathway intermediate **9** (might accumulate if *N*-oxygenation is the final biosynthetic step). One further intriguing observation was the fact that the unusual purple colour of WT production cultures is lost in both Δ *obaJ* (Figure 4.7a) and Δ *obaL* strains. This phenotype was rescued in both genetically and chemically complemented strains where production of **1** had successfully been restored. This observation was made for all other Δ *oba* strains in this chapter, and the significance of this colour change is discussed in further detail in Chapter 8.

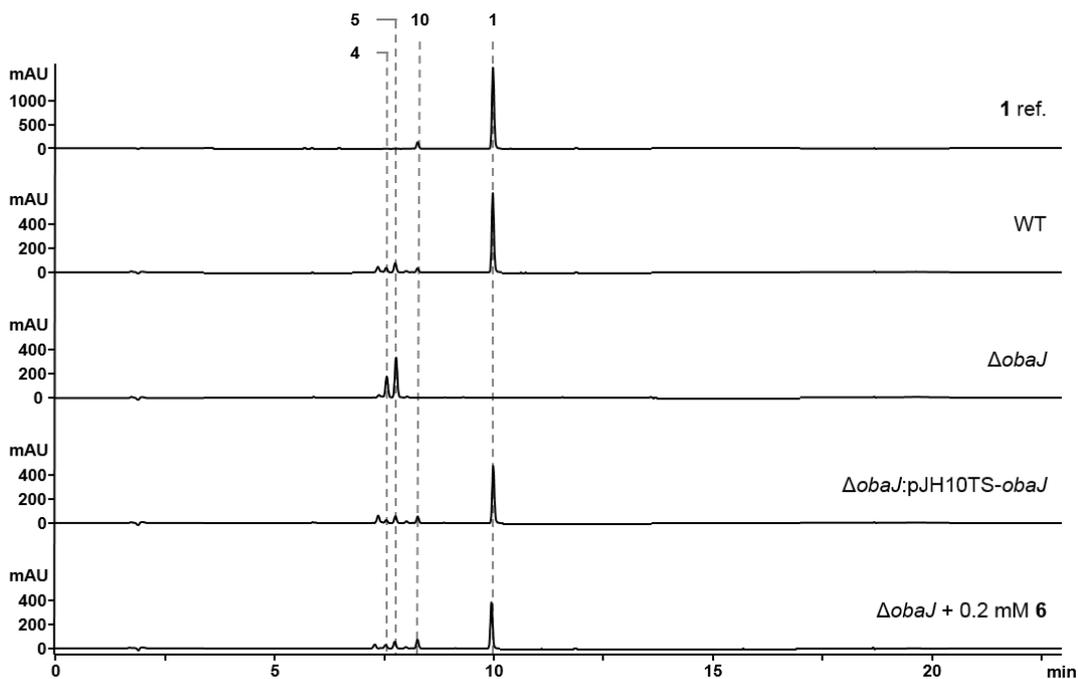


Figure 4.5. HPLC profiles for *obaJ* mutagenesis and complementation experiments at 270 nm. Numbered peaks refer to key products and shunt metabolites identified in Figure 4.4.

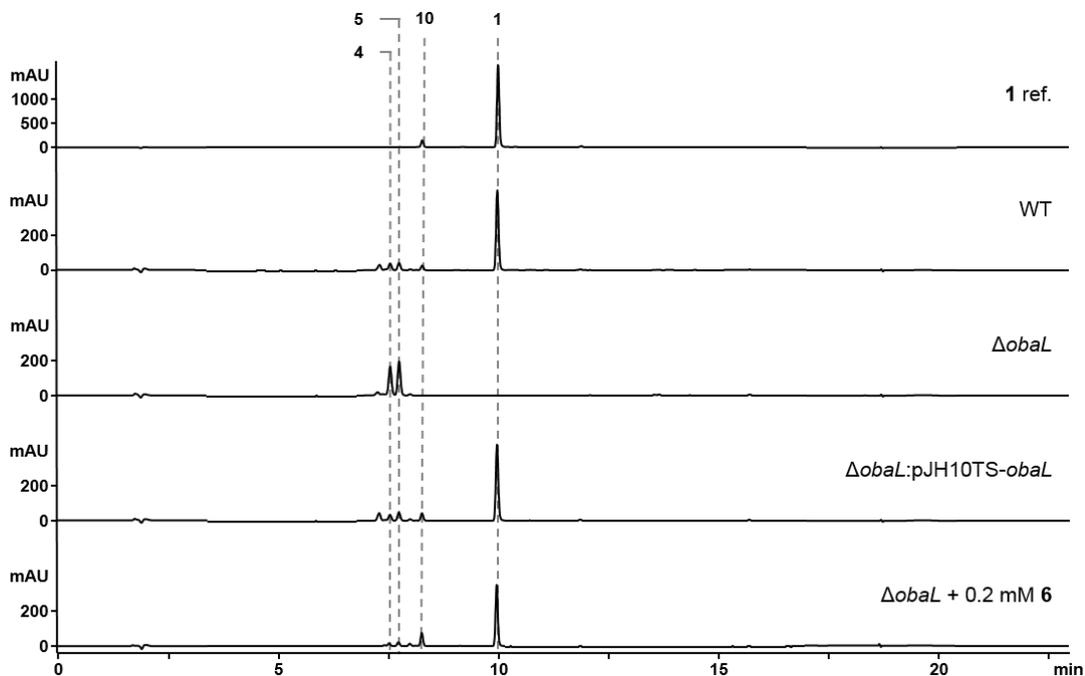


Figure 4.6. HPLC profiles for *obaL* mutagenesis and complementation experiments at 270 nm. Numbered peaks refer to key products and shunt metabolites identified in Figure 4.4.

4.4.2 Disrupting the biosynthesis of 4-APP (7)

obaDEF appear sufficient to encode for the production of **7** (Figures 3.7 and 4.4), the biosynthesis of which is well understood from previous genetic studies on chloramphenicol (He *et al.*, 2001) and pristinamycin (Blanc *et al.*, 1997) biosynthesis. Consistent with these predictions, deletion of the bifunctional gene *obaF* (a bifunctional 4-aminochorismate mutase/4-aminoprephenate dehydrogenase) led to the loss of **1**, **4** and **5** production (Figure 4.8), which further indicated that these two shunt metabolites originate from some element of **2** biosynthesis. Addition of exogenous **3** (1 mM final conc.) was sufficient to restore production of **1** to between 68%-71% of WT titres, as was the ectopic expression of *obaF* which restored production to between 73%-90% of WT titres. **3** is presumably converted to **7** by transamination *in vivo* to restore production. As with Δ *obaJ* and Δ *obaL*, the purple-brown colour associated with **1** production is lost in Δ *obaF*, but is restored by genetic and chemical complementation (Figure 4.7b). Surprisingly, no accumulation of the pathway intermediate **6** could be observed in the *obaF* mutant.

4.4.3 Disrupting the biosynthesis of AHNB (2)

Mutation of *obaH*, which encodes a putative ThDP-dependent pPDC thought to be responsible for the biosynthesis of **2** (Figures 3.1 and 3.5), led to the abolition of **1** production (Figure 4.9). **4**, **5** and **6** did not accumulate in the Δ *obaH* strain, mirroring the Δ *obaF* phenotype. Ectopic expression of *obaH* led to the recovery of **1** production, albeit at lower titres of around 29% of WT production levels. This is reflected in the incomplete regeneration of the WT production culture colour phenotype (Figure 4.7c). Unfortunately, the addition of exogenous synthetic **2** or **9** (0.2-1 mM final conc.), the anticipated products of ObaH, did not recover **1** production in these mutants. One possible explanation for this is simply that they fail to penetrate the *Pseudomonas* cell membrane. Whilst it is clear that ObaH is important for the biosynthesis of **1**, its exact function could not be determined by mutational analysis alone.

4.4.4 Disrupting obafluorin (1) dipeptide assembly

Deletion of *obaI*, which encodes a putative dimodular NRPS, abolished production of **1**, which was recovered to 79% of WT levels following ectopic complementation (Figure 4.10). The Δ *obaI* mutant accumulated the shunt metabolites **4** and **5** at similar levels to the WT strain, but, once again, exhaustive analysis of the fermentation extracts did not show any accumulation of **2** or **6**, the putative substrates of Obal. Deletion of the discrete ArCP *obaK* gave similar results to the Δ *obaI* mutant as expected, with **1** production abolished, and this was recovered by ectopic expression

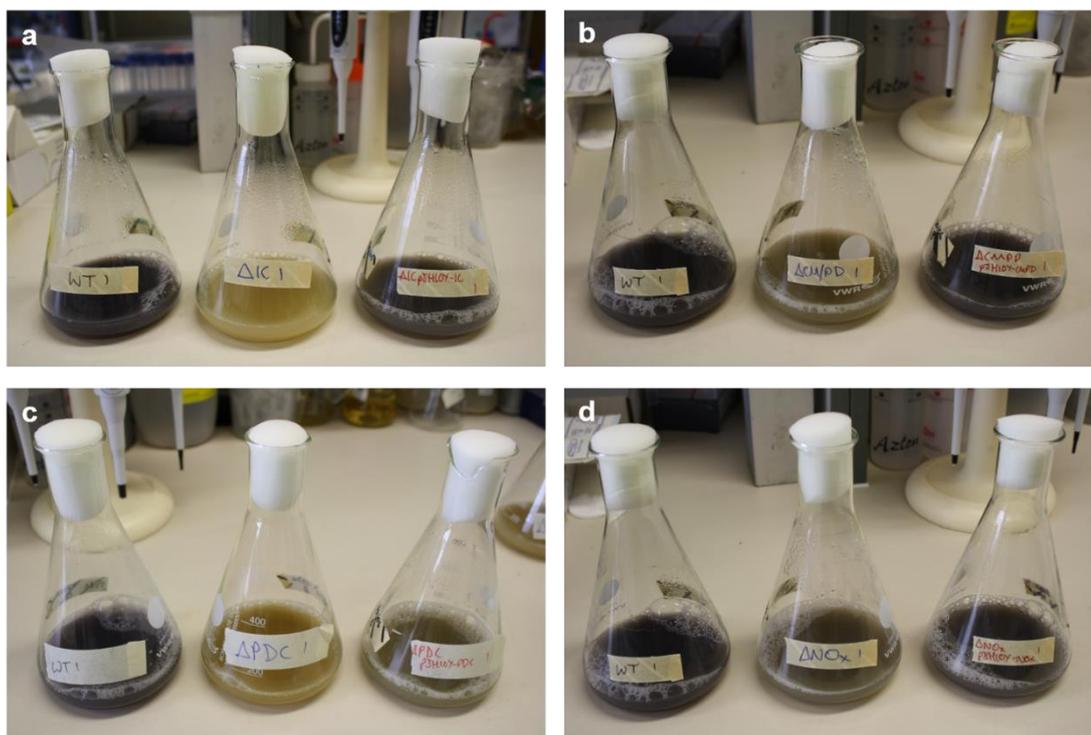


Figure 4.7. Production culture flasks for WT, Δ and genetically complemented Δ strains. Left = WT; Middle = Δ strain; and right = genetically complemented Δ strain. (a) Δ obaJ ('IC' = isochorismatase); (b) Δ obaF ('CM/PD' = chorismate mutase/prephenate dehydrogenase); (c) Δ obaH ('PDC' = phenylpyruvate decarboxylase); and (d) Δ obaC ('NOx' = N-oxygenase).

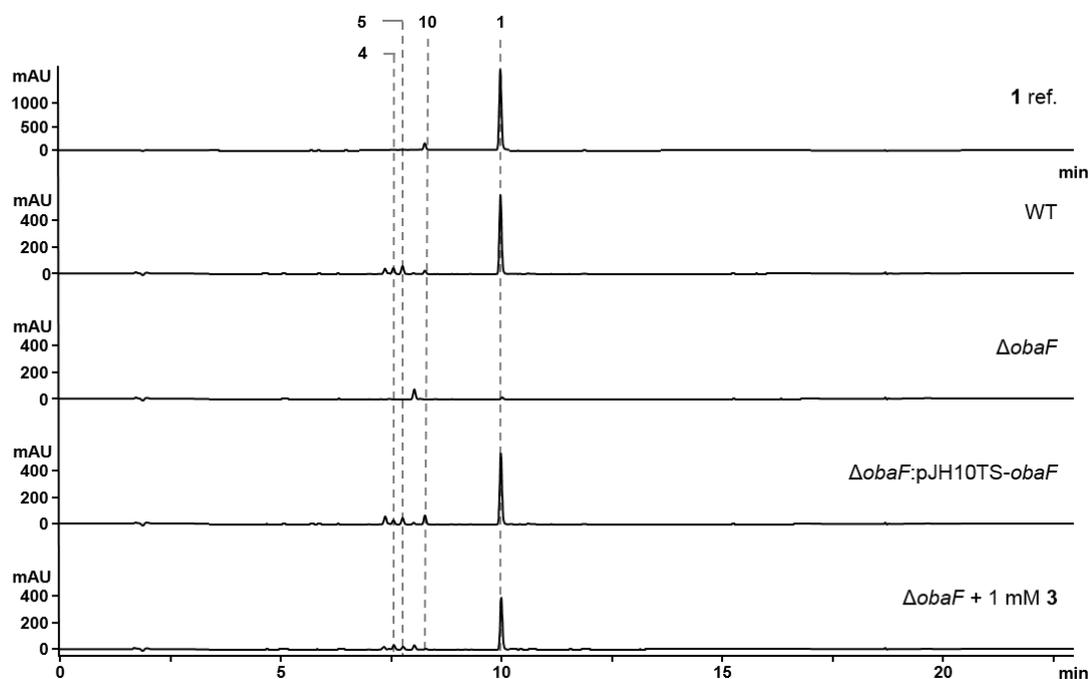


Figure 4.8. HPLC profiles for obaF mutagenesis and complementation experiments at 270 nm. Numbered peaks refer to key products and shunt metabolites identified in Figure 4.4.

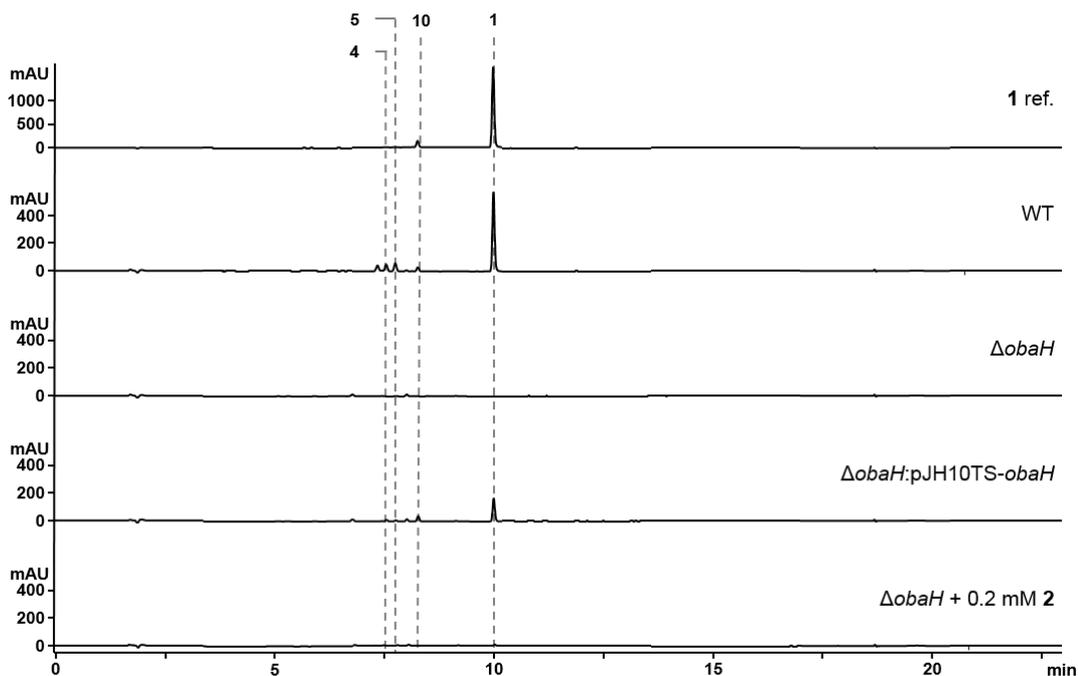


Figure 4.9. HPLC profiles for *obaH* mutagenesis and complementation experiments at 270 nm. Numbered peaks refer to key products and shunt metabolites identified in Figure 4.4.

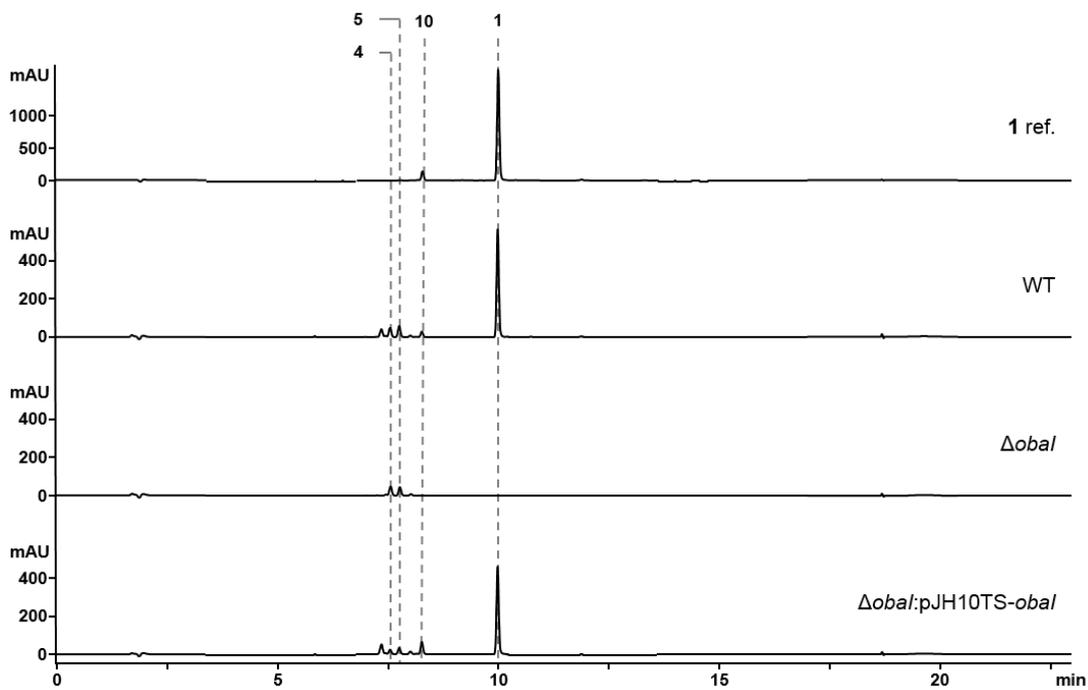


Figure 4.10. HPLC profiles for *obal* mutagenesis and complementation experiments at 270 nm. Numbered peaks refer to key products and shunt metabolites identified in Figure 4.4.

of *obaK* to 77% of the WT level (Figure 4.11). The titres of **4** and **5** also increased significantly (approx. four-fold) but again neither **2** nor **6** accumulated. Exogenous supply of **6** did not lead to production of **1** confirming that this phenotype is not linked to a lack of **6** biosynthesis. This possibility was queried because *obaK* appears to have evolved from the splitting of an ancestral *entB*-like gene, which usually encode for didomain proteins involved in the biosynthesis of **6** (Walsh *et al.*, 1990; Drake *et al.*, 2006). PCR and Sanger sequencing confirmed that *obaK* and *obaJ* represent independent PCSs. Both $\Delta obaI$ and $\Delta obaK$, and their complemented mutants, displayed identical phenotypes in production cultures to their equivalents for *obaJ* as illustrated in Figure 4.7a.

4.4.5 Disrupting nitro group formation

The gene *obaC* encodes a putative non-heme di-iron monooxygenase expected to oxidise the **1** aromatic amino group derived from **7**. This was predicted to utilise one of several pathway intermediates as its substrate (Figure 3.8c), but the precise substrate could not be identified by *in silico analysis* alone. Deletion of *obaC* abolished production of **1**, **4** and **5** (Figure 4.12), and the biosynthesis of all three compounds was re-established after ectopic complementation with *obaC* to 81% of WT titres of **1**. Surprisingly, only the addition of exogenous synthetic **8** (0.2 mM final conc.) to growing cultures of this mutant could restore the biosynthesis of **1** to 29% of WT titres, whereas addition of **2** did not. Furthermore, no shunt metabolites comprising an aminoaryl moiety were detected, consistent with ObaC functioning during precursor biosynthesis rather than as a final modification, likely preceding the activity of ObaH (Figure 4.4). *N*-oxygenase homologues AurF and CmlB, involved in aureothin and chloramphenicol biosynthesis respectively (He and Hertweck, 2001; Knoot *et al.*, 2016), also function at an earlier stage in biosynthesis (Figures 3.8a and b). The characteristic colour of WT cultures is lost in $\Delta obaC$ and is replaced by a similar murky green colour to the $\Delta obaF$ strain (Figure 4.7d), perhaps offering further coincidental evidence that ObaC functions previous to ObaH.

4.4.6 Disrupting precursor supply

obaM, encoding a putative type II DAHP synthase, was predicted to be involved in ensuring a stable supply of chorismate for the biosynthesis of **2** and **6**, mitigating the effects of negative feedback inhibition experienced by type I enzymes. However, whilst deletion of *obaM* led to the abolition of **1** production (Figure 4.13), it was surprising to observe the accumulation of **4** and **5** in this mutant, which indicated that the biosynthesis of **2** was proceeding normally. Moreover, exogenous supply of **6**

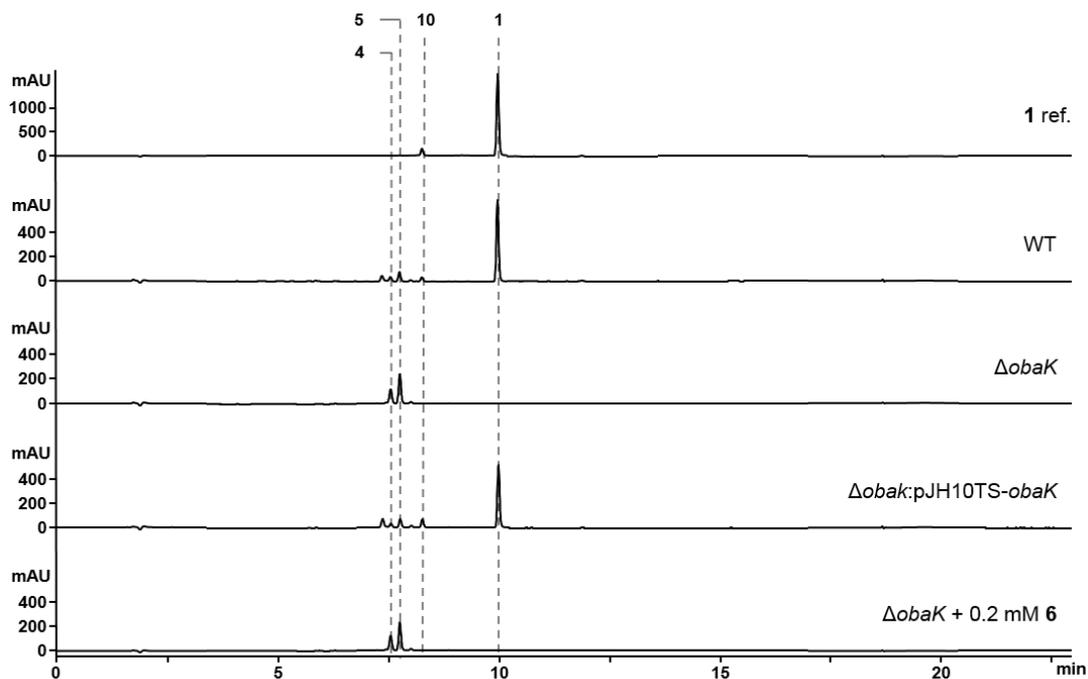


Figure 4.11. HPLC profiles for *obaK* mutagenesis and complementation experiments at 270 nm. Numbered peaks refer to key products and shunt metabolites identified in Figure 4.4.

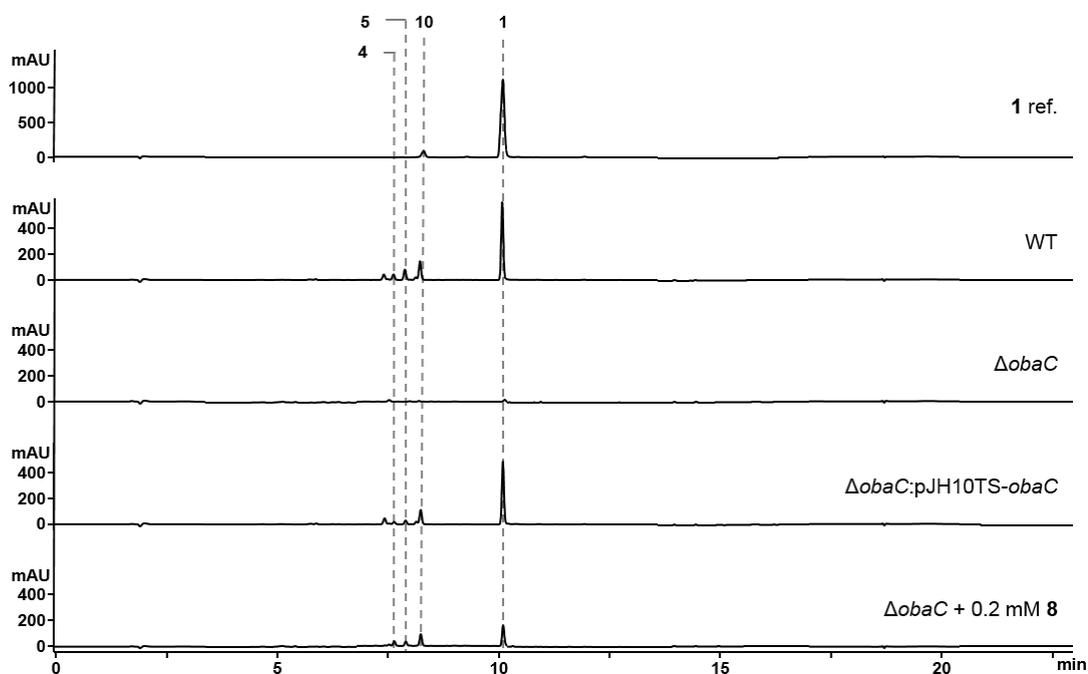


Figure 4.12. HPLC profiles for *obaC* mutagenesis and complementation experiments at 270 nm. Numbered peaks refer to key products and shunt metabolites identified in Figure 4.4.

(0.2 mM final conc.) alone was sufficient to restore production of **1** to 68–72% of WT titres, further indicating that biosynthesis of **2** is unperturbed by *obaM* deletion. Consistent with this, exogenous supply of **3** (transaminated to **7** - see 4.4.2) alone could not recover **1** production, but did lead to further elevated levels (five- to seven-fold) of **4** and **5**. Initial attempts to genetically complement this mutant by introducing pJH10TS-*obaM* failed, indicating possible polar effects of the mutation on genes involved in the biosynthesis of **6**. The putative *obaM* methionine start codon for this gene was selected based on Prodigal PCS identification and homologous sequences returned in blastn database searches, however further analysis of the genome sequence revealed three alternate potential start codons. Cloning of the gene using each possible start codon in pJH10TS and their introduction into the Δ *obaM* background allowed the identification of the true start codon – a valine (GTG) forty-one codons upstream of the predicted start site that overlaps with the PCS of *obaL*. The PCSs of *obaJKLMN* thus overlap to form a putative operon, perhaps reflecting the observed dependence of enzymes involved in **6** biosynthesis on ObaM function in chorismate supply. Ectopic expression of *obaM* using this alternate start codon recovered **1** production to 50–55% of WT titres, precluding the possibility of polar effects on *obaJLN*. For Δ *obaM* strains, the purple colour is lost in the mutant strain but is restored by genetic or chemical complementation (i.e. not when **3** is fed alone to Δ *obaM*), creating similar phenotypes as illustrated for Δ *obaJ* (Figure 4.7a).

Initially we proposed that ObaG, a putative SHMT, in conjunction with the GCS encoded by *orfs2345* (Figures 3.1 and 3.10), could function to supply glycine to drive flux through the ObaH-catalysed ThDP-dependent reaction to form **2**. Given the pool of free glycine in the cell supplied via primary metabolism, one might expect a drop in **1** titre as a result of disrupting either the *orf2345*-encoded GCS or ObaG. However, a deletion across *orf2345* did not result in an observable decrease in **1** titre (or loss of purple colour), which could be explained by the richness of **1** production medium (OPM) obviating the need for additional mechanisms of glycine supply. Surprisingly, disruption of *obaG* resulted in the complete loss of **1** production, and the accumulation of **4** and **5** (Figure 4.14), as well as loss of purple colour from production cultures (Δ *obaJ* phenotype – Figure 4.3a). This indicated a potential biosynthetic role for ObaG; the accumulation of shunt metabolites **4** and **5** suggested their origin from an intermediate between ObaH and ObaG, placing ObaG as the enzyme directly responsible for the biosynthesis of **2**. This strongly suggested that 4-nitrophenylacetaldehyde (4-NPA; **11**) might be the key intermediate, and Hebert and Knaggs had also previously suggested that the amino equivalent

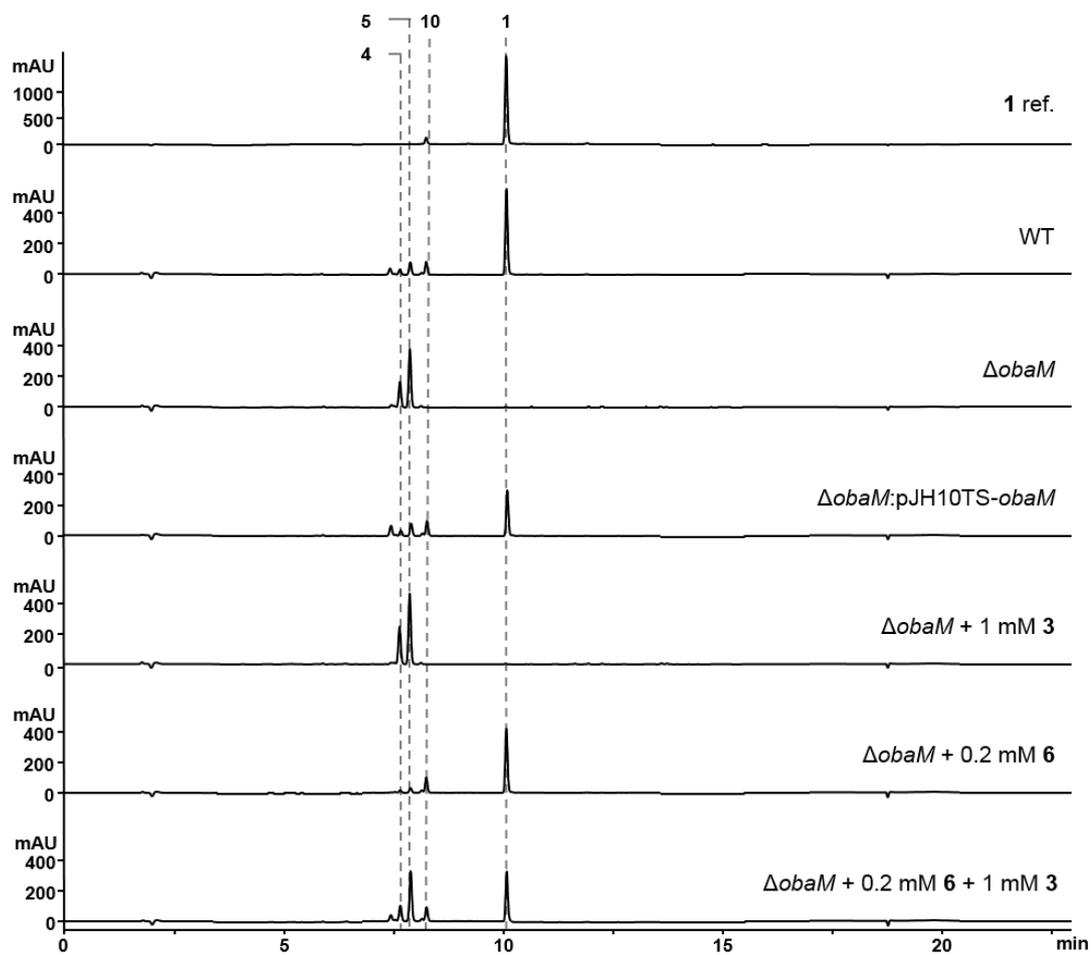


Figure 4.13. HPLC profiles for *obaM* mutagenesis and complementation experiments at 270 nm. Numbered peaks refer to key products and shunt metabolites identified in Figure 4.4.

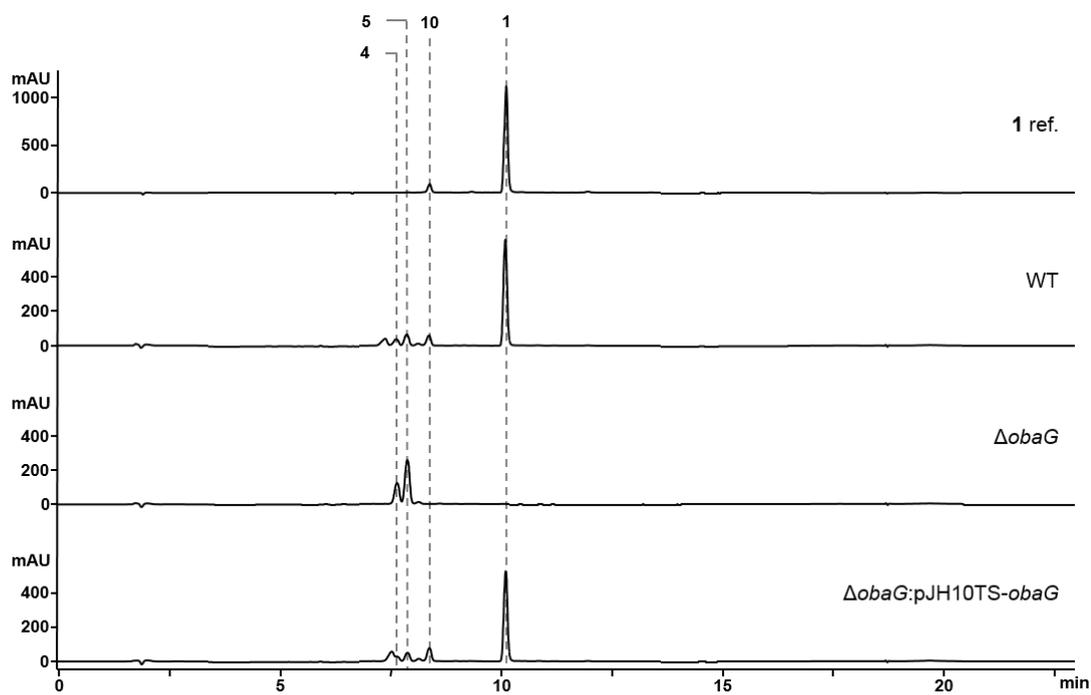


Figure 4.14. HPLC profiles for *obaG* mutagenesis and complementation experiments at 270 nm. Numbered peaks refer to key products and shunt metabolites identified in Figure 4.4.

4-aminophenylacetaldehyde, might be an alternative precursor in the biosynthesis of **2** (Herbert and Knaggs, 1990; 1992b). Consistent with this hypothesis, the addition of exogenous **11** (0.2 mM final conc.) to growing cultures of the $\Delta obaG$ mutant did not restore **1** production, but did lead to elevated levels of **4** and **5**. Ectopic expression of *obaG* in the deletion strain restored **1** production to 62% of WT titres (and purple culture colour as for *obaJ* – Figure 4.7a) but again, **2** and **9** failed to recover **1** production as before, necessitating further experiments to ascertain the exact function of ObaG.

4.5 Discussion

Acquisition of the sequence data for the *oba* BGC allowed *in vivo* mutational experiments to be designed and performed. To this end, the suicide vector pTS1 was developed to create in-frame deletions in a number of proposed *oba* biosynthetic genes. To ensure individual deletions did not result in polar effects on other biosynthetic genes, the complementation vector pJH10TS was also developed for the ectopic expression of WT genes in their respective mutant strain backgrounds.

All knockouts of proposed biosynthetic enzymes led to the abolition of **1** production, confirming that the identified BGC is responsible for the biosynthesis of **1**. Furthermore, all genes disrupted could be complemented genetically, giving confidence that any observed alterations in metabolite production profiles were the result of an individual specific PCS disruption and not due to polar effects.

Aside from the abolition of **1** production, the only other noticeable changes in metabolite profiles for Δoba strains was the accumulation of **4** and **5**. These shunt metabolites were initially determined to originate from some intermediate in the biosynthesis of **2**, given that they failed to accumulate in $\Delta obaC$, $\Delta obaF$ and $\Delta obaH$ backgrounds. This result supported the hypothesis that ObaF was functioning analogously to CmlDC and PapBC in chloramphenicol and pristinamycin biosynthesis respectively, and was further supported by the return of **1** production with exogenous supply of **3**. By association, it is reasonable to assume that ObaD and ObaE also function analogously to homologues in those pathways to convert chorismate to **7** (Figure 3.7a). At no point were any metabolites comprising an aminoaryl group detected, indicating that ObaC functions early during the biosynthesis of **2**, probably before the activity of ObaH (Figure 4.4). This was supported by the absence of

nitroaryl shunt metabolites **4** and **5** in $\Delta obaH$ production cultures, and the fact that addition of exogenous **8** restored **1** production in the $\Delta obaC$ mutant.

ObaJLN are also assumed to function analogously to their homologues in enterobactin biosynthesis (Figure 3.6a), given that **1** production is abolished with concurrent accumulation of **4** and **5** in both $\Delta obaJ$ and $\Delta obaL$. Recovery of production by addition of exogenous **6** indicated that it is indeed the product of these three enzymes. The suggested role in **1** assembly for the NRPS ObaI and ArCP ObaK was similarly supported by the abolition of **1** production and accumulation of **2**-derived nitroaryl shunt metabolites in their respective mutant backgrounds. Given that *obaK* appears to have evolved from the splitting of an ancestral *entB*-like gene (didomain ArCP/isochorismatase), **6** was also exogenously fed to this mutant but did not result in **1** production, indicating that it represents a discrete enzyme that is not required for the biosynthesis of **6**.

One slightly confounding result was the finding that an $\Delta obaM$ deletion only impacted the biosynthesis of **6** – evidenced by the fact that both **4** and **5** accumulate in this mutant, which suggested that enzymes responsible for the biosynthesis of **2** are still functional. This observation was supported by the recovery of **1** production following exogenous supply of **6** alone, and the failure to restore production on addition of **3** alone (which is readily transaminated to **7** in the cell, thereby supplying an intermediate in the biosynthesis of **2** – see 4.4.2). This is surprising because the proposed role of ObaM is in ensuring sufficient chorismate is supplied to the pathway which is required for the biosynthesis of both **1** precursor acid substrates (Figures 3.10 and 4.4). The apparent impact on the biosynthesis of **6** is not completely without precedent however as a DAHP synthase in the vanchrobactin BGC was shown to be essential for the growth of *Vibrio anguillarum* under iron-limiting conditions (Balado *et al.*, 2008). Vanchrobactin comprises a single **6**-derived moiety, though the impact of type II DAHP synthase deletion on the biosynthesis of **6** in this case is also not understood. Furthermore, *obaM* is located in an operon with *obaJKLN*, perhaps reflecting a dependence of their products on the presence of ObaM. The differential effects of the $\Delta obaM$ mutation on the biosynthesis of **1** precursors are hard to rationalise but possible explanations might include the necessity for protein-protein interaction between ObaM and one or more of ObaJLN, or that the biosynthesis of **7** can occur via an alternate mechanism to that proposed here for ObaDEF, perhaps from conversion of a metabolite downstream from chorismate. The latter suggestion is more likely given that homologues of *obaJLN* occur in BGCs that do not comprise

a type II DAHP synthase, e.g. enterobactin (Walsh *et al.*, 1990). Deletions in *obaD* and *obaE* should be performed to confirm that these enzymes fulfil identical roles to their homologues in chloramphenicol and pristinamycin biosynthesis (Blanc *et al.*, 1997; He *et al.*, 2001), and that **7** cannot be supplied via an alternate route.

This left the biosynthesis of **2** as the remaining unknown, and the product of *obaH* to be functionally assigned. Its deletion led to the abolition of **1** production and nitroaryl shunt metabolites did not accumulate, implicating the involvement of ObaH in the biosynthesis of **2**. Unfortunately, whilst genetic complementation was possible, the introduction of either synthetic **2** or **9** failed to restore **1** production. This is predicted to be due to the inability of these substrates to cross the *P. fluorescens* cell membrane and mutagenesis alone was thus not sufficient to demonstrate the exact function of ObaH in the biosynthesis of **2**. However, the disruption of *obaG* also led to the abolition of **1** production, indicating an unforeseen biosynthetic role. The accumulation of both **4** and **5** in Δ *obaG* but not Δ *obaH* suggested that ObaG performs an additional step between ObaH and ObaK-mediated assembly, and that **11** might be the key intermediate. **11** was confirmed as the source of **4** and **5** based on elevated levels of their accumulation when **11** was exogenously fed to Δ *obaG*. Feeding with **2** or **9** failed to restore **1** production in the Δ *obaG* background, as for ObaH, and so again further experimental work was required to characterise the ObaG-catalysed reaction. This is the subject of Chapter 5.

These observations led to the formulation of an alternative biosynthetic scheme for **2** (Figure 4.15) in which ObaH is a ThDP-dependent decarboxylase that generates **11** by decarboxylating **8** as a substrate for ObaG. ObaG might then act as a transaldolase rather than an SHMT, catalysing the retro-aldol cleavage of L-serine to generate a PLP-stabilised quinonoid intermediate and formaldehyde, before subsequently catalysing C-C bond formation between enzyme-bound glycine and **11** to form **2**. This hypothetical pathway has three advantages over an entirely ThDP-dependent mechanism: 1) The use of **11** as an intermediate explains the absence of **4** and **5** accumulation in Δ *obaC*, *F* and *H* mutants but their presence in the Δ *obaG* background; 2) the use of an amino acid substrate obviates the need for a transaminase (not present in the cluster) to install the α -amino group on **2**; and 3) The C₁ formaldehyde unit by-product of the reaction could then be recycled by the *orf2345*-encoded GCS, to generate further L-serine from glycine and drive flux through the ObaG-catalysed reaction. This system was initially proposed to be supplying glycine

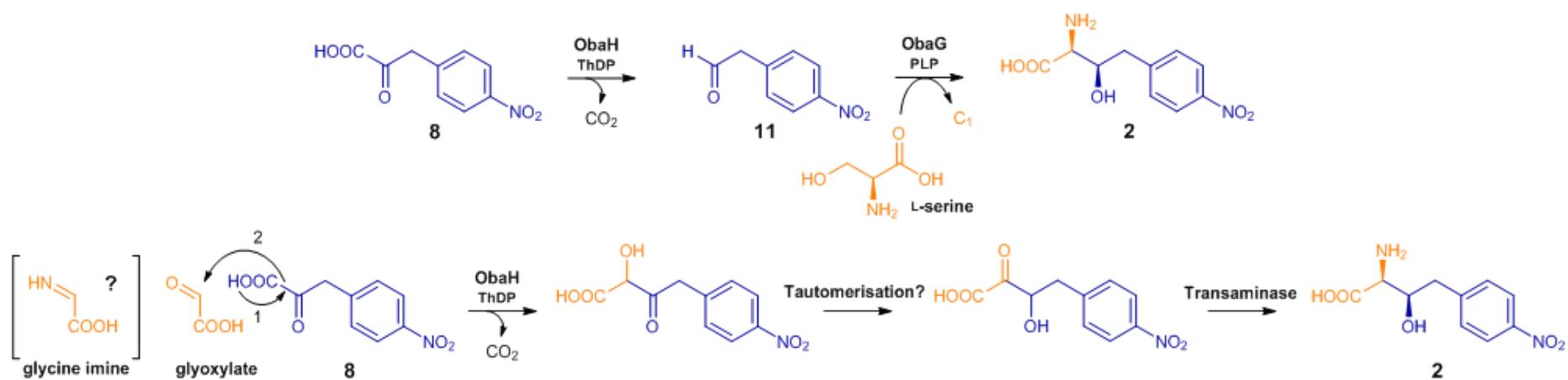


Figure 4.15. First proposed biosynthetic mechanism for the biosynthesis of AHNB (2) involving ObaG acting as a transaldolase. The initial hypothesis proposed by Herbert and Knaggs (1990) is re-illustrated below. Glycine imine (boxed) could also be a substrate for ObaH – discussed in greater detail in Chapter 5.

(glyoxylate) to the ThDP-dependent reaction but could be operating in the reverse direction to generate serine as a substrate for a PLP-dependent mechanism (Figure 3.10). Disruption across these four genes did not lead to **1** production loss and further experiments based on minimal conditions would be required to determine the role of *orf2345* in the biosynthesis of **1**, if any. Another possibility is that ObaG is an SHMT/L-TA that has evolved to efficiently catalyse an aldol-type C-C bond formation reaction between glycine and **11**, without initial amino acid cleavage. However, this seems unlikely given previous attempts to engineer these enzymes to catalyse aldol reactions (Fesko, 2016).

The accumulation of pathway intermediates and shunt metabolites in *oba* locus mutants was limited to the breakdown products of **11**. At no stage were any aminoaryl breakdown products, **2** or **6** detected, despite exhaustive searches for them. The former can be explained by the functioning of the *N*-oxygenase ObaC at an early stage in biosynthesis and the substrate specificities of subsequent enzymes requiring the presence of the nitroaryl group. Absence of **2** from $\Delta obaJ$, *L*, *K* and *I* strains is likely due to the breakdown of this nonproteinogenic amino acid for recycling through primary metabolic pathways. Indeed, the L-*threo* stereochemistry of **2** (see Chapter 5) would potentially make it a substrate for cleavage by an L-threonine aldolase (L-TA), a primary metabolic enzyme that is able to cleave β -OH- α -AAs to generate glycine and an aldehyde by-product (Fesko, 2016). In this case the aldehyde product would be **11**, which has been shown to be degraded to **4** and **5**. The absence of **6** can be explained by the observation that cultures of $\Delta obaJ$ and $\Delta obaL$ supplied exogenous **6** at concentrations above ~0.4 mM, both growth of *P. fluorescens* and biosynthesis of **1** are adversely affected, and thus it is likely prevented from accumulating in the cell. Taken together these combined data suggest that the biosynthetic pathway is tightly regulated to avoid the accumulation of intermediates deleterious to cell growth and survival. Strategies to avoid the accumulation of toxic intermediates in NP biosynthetic pathways are well reported. A recent example relevant to this work is found in ripostatin biosynthesis, in which the intermediary of free phenylacetaldehyde, a reactive aldehyde structurally similar to **11**, is avoided via the activity of a pyruvate dehydrogenase-like complex. This complex functions to decarboxylate phenylpyruvate and then retain the phenylacetaldehyde product as a phenylacetyl-S-carrier protein species (Fu *et al.*, 2017). Whilst not necessarily conducive to the elaboration of steps in a biosynthetic pathway, mechanisms of toxic build up reflect the elegant, tight regulation and coordination that has evolved in NP biosynthesis.

Finally, whilst the lysine tRNA gene represents one likely boundary of the *oba* BGC, the boundary upstream of *obaA* remains to be defined. Possible roles for *orfs12345678* were proposed in Chapter 3, and a simple experiment to determine their involvement in the biosynthesis of **1** would be to perform a large deletion across all 8 PCSs. If **1** production is affected, then subsequent smaller deletions can be performed to identify individual PCSs responsible. The PK cyclase encoded in the cluster has also not been studied, though this is likely to be an artefact of HGT that is inherited with the cluster by virtue of its position between **1** biosynthetic loci and the lysine tRNA gene. pTS1-mediated disruption of this gene should determine whether it is significant for the biosynthesis of **1** or not, but no biosynthetic role could be ascribed based on *in silico* analysis. The threonyl-tRNA synthetase (ThrRS) gene located between *obaN* and the lysine tRNA gene will be explored further in Chapter 8.

The mutagenesis and complementation experiments described here have confirmed the identity of the BGC necessary for **1** production and this was deposited in GenBank under the accession no. KX931446. Furthermore, the likely function of *oba* loci was determined, in addition to gaining further understanding of the order of biosynthetic steps (e.g. *N*-oxygenation). The absence of toxic precursors/shunt metabolites in Δ *oba* strains, and the importance of *obaM* in the biosynthesis of **6**, indicates that the **1** biosynthetic pathway is subject to very tight regulation. Critically, whilst these experiments demonstrated the essentiality of *obaH* in **1** production, and identified that a potential biosynthetic role for *obaG*, failed chemical feeding experiments with **2** and **9** left the exact function of these enzymes during the biosynthesis of **2** to be determined. In order to achieve this, an alternative approach described earlier, *in vitro* reconstitution, was applied to characterise the function of these two enzymes independently of the cell and potential interference from primary metabolic pathways. These experiments are described in Chapter 5.

Chapter 5:
Biochemical characterisation
of AHNB (**2**) biosynthesis

Chapter 5: Biochemical characterisation of AHNB (2) biosynthesis

5.1 Introduction

Mutational analysis of *oba* biosynthetic loci revealed that the correct BGC for the biosynthesis of obafluorin (**1**) had been identified and, critically, that *obaH*, encoding a ThDP-dependent pPDC, is essential for its production. Furthermore, the absence of 4-NPE (**4**) and 4-NPAA (**5**) accumulation in the Δ *obaH* strain was consistent with a role in the biosynthesis of AHNB (**2**). **8** appeared to be a likely substrate of ObaH, which was proposed to introduce two further carbons to the side chain from glyoxylate, though this was not validated experimentally. ThDP-dependent enzymes are ubiquitous in nature and participate in numerous biosynthetic pathways, catalysing a wide range of reactions typically involving formation or cleavage of C-C bonds, making one an ideal candidate for this reaction (Jordan, 2003; Müller *et al.*, 2013). The first step common to all ThDP-dependent reactions is the formation of a ThDP ylide, resulting from protein-assisted deprotonation at its C-2 atom (Andrews and McLeish, 2013), typically mediated by a highly-conserved glutamic acid active site residue. This activates the cofactor, allowing the subsequent decarboxylation of an α -keto acid to generate an active aldehyde 'acceptor' intermediate with two free electrons to participate in subsequent C-C bond formation.

In the context of **2** biosynthesis, Herbert and Knaggs (1990) proposed that either **7** or **8** act as the α -keto acid donor that is decarboxylated for subsequent C-C bond formation with glyoxylate to generate an α -hydroxy- β -keto acyloin intermediate (Figures 1.22 and 3.1). This would then need to be tautomerised to form the β -hydroxy- α -keto equivalent for subsequent transamination to **2**. Tautomerisation could occur either enzymatically or spontaneously, and if a subsequent transaminase was specific for the β -hydroxy- α -keto substrate, it would represent the ideal mechanism to shift the equilibrium of the reaction in favour of the necessary tautomer for **2** formation by kinetic resolution. By this logic, a transaminase specific for the α -hydroxy- β -keto acyloin would create the opportunity to access α -hydroxy- β -amino acids and the enhanced proteolytic stability they afford, as alluded to in the introduction. ThDP-dependent enzymes are known to accept an extremely diverse set of substrates, and an alternate route to **2** using glycine imine as a substrate was also considered (Figure 4.15). In this case, neither tautomerisation nor transamination would be required to generate **2**. However, this is unlikely given stable isotope feeding experiments by

Herbert and Knaggs that showed glycine to be an excellent donor for C-1 and C-2 in **2** (Herbert and Knaggs, 1990). There is also no natural ThDP-dependent enzyme reported to date that uses an imine substrate, though this does not preclude the possibility that such an enzyme could evolve in a specialised metabolic context. The broad use of α -keto substrates by the ThDP-dependent enzyme family makes the scheme proposed by Herbert and Knaggs the most likely for a purely ThDP-mediated mechanism.

Unexpectedly however, *obaG* was also implicated in the biosynthesis of **2**, leading to abolition of **1** with concomitant accumulation of **4** and **5** when disrupted genetically. This led to the proposition of an alternative scheme (Figure 4.15) for **2** biosynthesis that proceeds via a PLP-dependent mechanism. PLP-dependent enzymes also catalyse an extremely broad set of reactions including transaminations, decarboxylations and β -eliminations among others, almost all involving amino compound substrates (Schneider *et al.*, 2000). PLP is very efficient at resonance stabilising α -carbanionic intermediates, formed by the removal of a proton, carboxylate or side chain from the substrate C_{α} atom. The resulting quinonoids are susceptible to attack by electrophilic substrates, resulting in the formation of new C-C bonds with the C_{α} . The product of *obaG* was initially annotated as a putative SHMT, enzymes which, in the absence of THF, deprotonate L-serine to release formaldehyde and generate a glycine-bound quinonoid intermediate (Figure 1.11 - Schirch *et al.*, 1985). Glycine is subsequently released by hydrolysis. Based on mutagenetic data, a mechanism was proposed for ObaG by which it acts as a transaldolase, first catalysing L-serine cleavage to yield glycine, before performing a C-C bond forming reaction with **11** to release **2** (Figure 4.15). L-Threonine is also a possible substrate if ObaG is acting as an L-threonine transaldolase (L-TTA). The L-TTAs are a rare family of fold-type I PLP-dependent enzymes that perform an identical cleavage reaction to L-TAs, but which subsequently catalyse C-C bond formation between glycine and an aldehyde substrate (Murphy *et al.*, 2001; Barnard-Britson *et al.*, 2012). A final possibility is that ObaG is catalysing the aldol-type reaction with **11** and glycine to form **2**, but this seems unlikely based on previous failed attempts to date to engineer SHMTs and L-TAs to perform such aldol-type reactions, where products are often generated in low yields and as a mixture of diastereoisomers (Fesko, 2016). However, one cannot rule out the possibility that an enzyme might evolve to perform this reaction in a stereospecific manner in a specialised metabolite context.

The failure of feeding experiments with either **2** or AHAB (**9**) left the exact roles of ObaH and ObaG in the biosynthesis of **1** to be defined. In this Chapter, the results of HPLC-based biochemical experiments are reported that dissect the individual and combined functions of ObaH and ObaG in the biosynthesis of **2**. Specifically, ObaG is identified as the enzyme directly responsible for **2** production and further experiments are described to evaluate the kinetics of this enzyme and its utility in synthetic approaches for the preparation of new β -OH- α -AAs.

5.2 Biochemical analysis of His₆-ObaH activity

ObaH was readily expressed as a soluble protein in a hexahistidine-tagged form in *E. coli* NiCo21(DE3):pLysS (Figure 5.1). Due to the original proposal of Herbert and Knaggs (1990) that a ThDP-dependent enzyme might produce an acyloin intermediate (Figure 3.1), ObaH was first assayed with **8** and glyoxylate as substrates (with ThDP and Mg²⁺ as cofactors) in a discontinuous format coupled to independent LCMS/MS and HPLC-UV assays (Materials and methods section 2.7.1). In none of my experiments was I able to detect the accumulation of an acyloin product, whereas **11** accumulated whenever His₆-ObaH was incubated with **8**, clearly identifying it as an **8** decarboxylase (Figure 5.2a). Synthetic references were used for comparison. Trial experiments were attempted with phenylpyruvate instead of **8**, but no decarboxylation product was detected, indicating the potential significance of the nitro group for ObaH substrate recognition. Unfortunately, the aminoaryl equivalent **7** is not commercially available and has not yet been synthesised in the lab to see whether it is accepted as a substrate by ObaH.

5.3 Biochemical analysis of His₆-ObaG activity

To examine the involvement of the putative SHMT-like enzyme ObaG in **2** biosynthesis, it too was expressed and purified as an N-terminally hexahistidine-tagged protein (Figure 5.1). A similar assay format as for His₆-ObaH was applied but using PLP rather than ThDP as cofactor, and using glycine and **11** as substrates (Materials and methods section 2.7.2), to first explore whether ObaG may catalyse an aldol-type reaction analogous to that of L-TAs in the formation of β -OH- α -AAs from glycine and an aldehyde. However, the production of **2** under these conditions was not observed. The experiment was repeated using either L-serine or L-threonine as the amino acid substrate to see if ObaG could function as a transaldolase. **2** production was only observed when L-threonine and **11** were present in the reaction mixture (Figure 5.2b), indicating that ObaG could indeed be a genuine L-TTA.

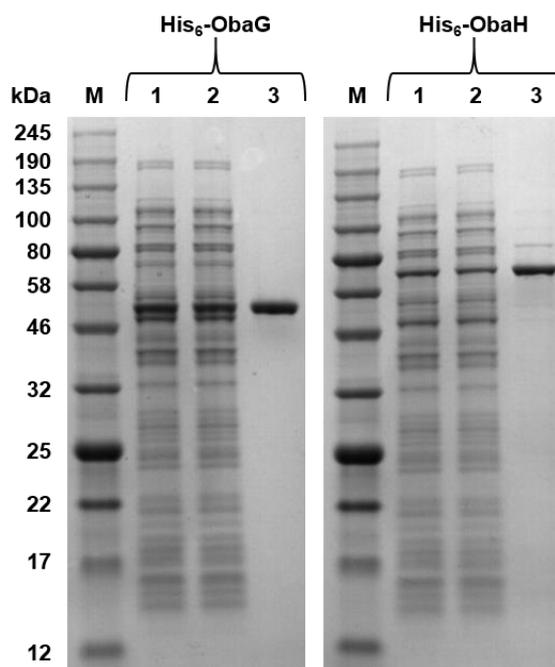


Figure 5.1. 12 % SDS PAGE gel of His₆-ObaG and His₆-ObaH purification steps. 1 = Cell lysate, 2 = Chitin wash eluent, 3 = Ni-affinity purified sample, and M = Colour Prestained Protein Standard, Broad Range (NEB). The expected molecular weights for His₆-ObaG and H are 50.8 and 64.2 kDa respectively.

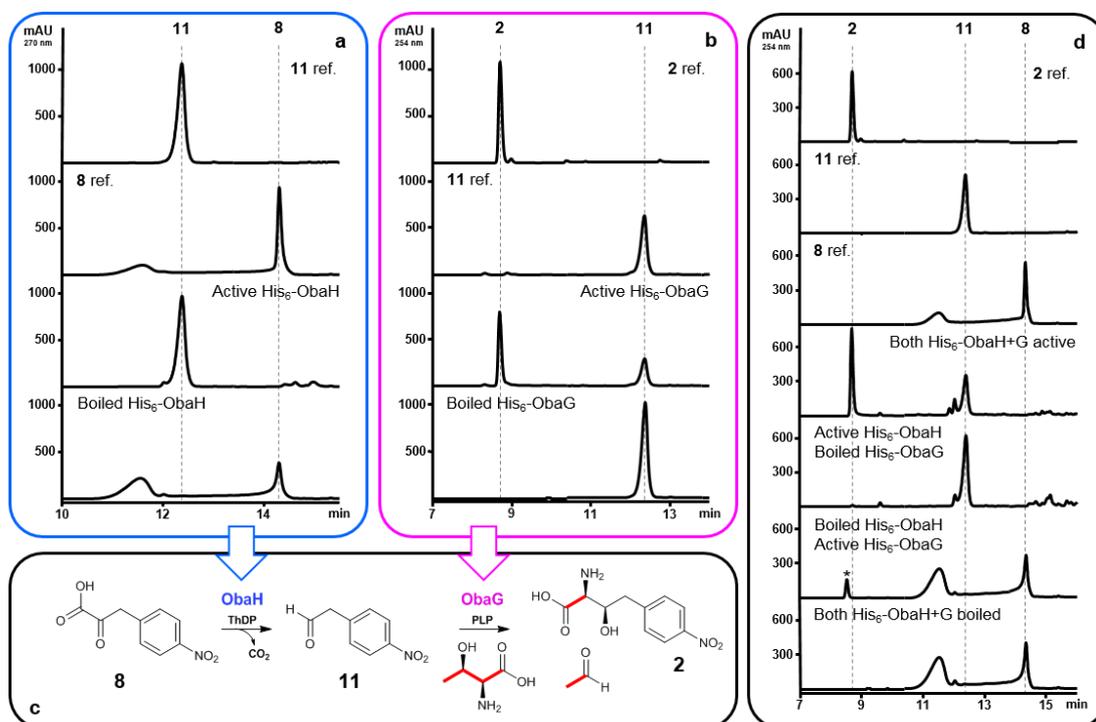


Figure 5.2. Experimental characterisation of the biosynthesis of AHNB (2). (a) His₆-ObaH decarboxylase activity assay HPLC profiles. (b) His₆-ObaG transaldolase activity assay HPLC profiles. (c) Illustration of the biosynthesis of 2 in a two-step reaction catalysed by consecutive action of ObaH and G. The His₆-ObaG transaldolase activity assay was also performed using [U-¹³C₄, ¹⁵N]L-threonine, and terminated reactions were purified and analysed by NMR (Figure 5.3) to show ¹³C incorporation into 2. The result of this experiment is also illustrated here. (d) Coupled decarboxylase and transaldolase assay HPLC profiles. 8 was also found to be a poor substrate for the ObaG-catalysed L-TTA reaction, yielding a new product *.

In order to confirm the L-TTA activity of ObaG, the biochemical assay described was repeated but using [U-¹³C₄,¹⁵N₁]L-threonine as a co-substrate with **11**. Dr. Daniel Heine analysed the samples (following C18 solid-phase extraction clean-up) using ¹³C NMR spectroscopy and HR-LCMS/MS, and was able to demonstrate the regiospecific transfer of three heavy isotope atoms to the appropriate positions in the product **2** (Figures 5.2c and 5.3), with doublets observed for the C-1 and C-2 positions sharing a coupling constant of $J_{C,C} = 52.8$ Hz. The coupling constant for the ¹³C and ¹⁵N atoms at the C-2 position was also determined ($J_{C,N} = 6.2$ Hz), indicating the ¹⁵N atom from [U-¹³C₄,¹⁵N₁]L-threonine was incorporated. The two acetal forms of glycerol and acetaldehyde ($J_{C,C} = 45.6$ Hz) were also detected, which indicated the formation of [1,2-¹³C₂]acetaldehyde occurred, identifying acetaldehyde as a by-product of the ObaG-catalysed reaction. The structure and stereochemistry (*L-threo*) of isolated **2** was verified by comparison of its NMR spectra and optical rotation to those of previously synthesised material (Pu *et al.*, 1994).

To confirm the biosynthetic relevance of L-threonine *in vivo*, [U-¹³C₄,¹⁵H₁]L-threonine was added exogenously to WT production cultures of *P. fluorescens* ATCC 39502 to a final concentration of 2 mM. This resulted in the highly efficient site-specific incorporation of a ¹³C₂¹⁵N₁ unit as anticipated, detected as a shift of 3 *m/z* using HR-LCMS/MS in the final **1** product from *m/z* [M+H]⁺ = 359.0875 (Expected *m/z* for C₁₇H₁₄N₂O₇: 359.0874, Diff. (ppm) = 0.28) to 362.0913 (Expected *m/z* for C₁₅¹³C₂H₁₄N¹⁵NO₇: 362.0911, Diff. (ppm) = 0.55), compared to no change in a WT culture fed unlabelled L-threonine at the same concentration (Figure 5.4). Finally, again using the same assay format (using 20 mM of starting substrates), ObaG was shown to also catalyse the reverse transaldol reaction to generate **11** and L-threonine from **2** and acetaldehyde as substrates (Figure 5.5), however yields of **11** were particularly low.

5.4 His₆-ObaH/His₆-ObaG coupled reaction experiment

As a final experiment to confirm the precise involvement of ObaH and G in the biosynthesis of **2**, both enzymes were incubated at equimolar concentrations with **8** and L-threonine (Materials and methods section 2.7.3). Excess PLP was removed from His₆-ObaG samples prior to the experiment to avoid unwanted side reactions with **8**. Following 2 h incubation, both **11** and **2** could be detected in an active enzyme sample, demonstrating that ObaH decarboxylase and subsequent ObaG L-TTA activities are sufficient to produce **2** (Figure 5.2c and d). As negative controls, reactions in which either one or both of His₆-ObaH and His₆-ObaG were boiled were performed. ‘Both’ boiled and ‘His₆-ObaG only’ boiled controls gave the anticipated

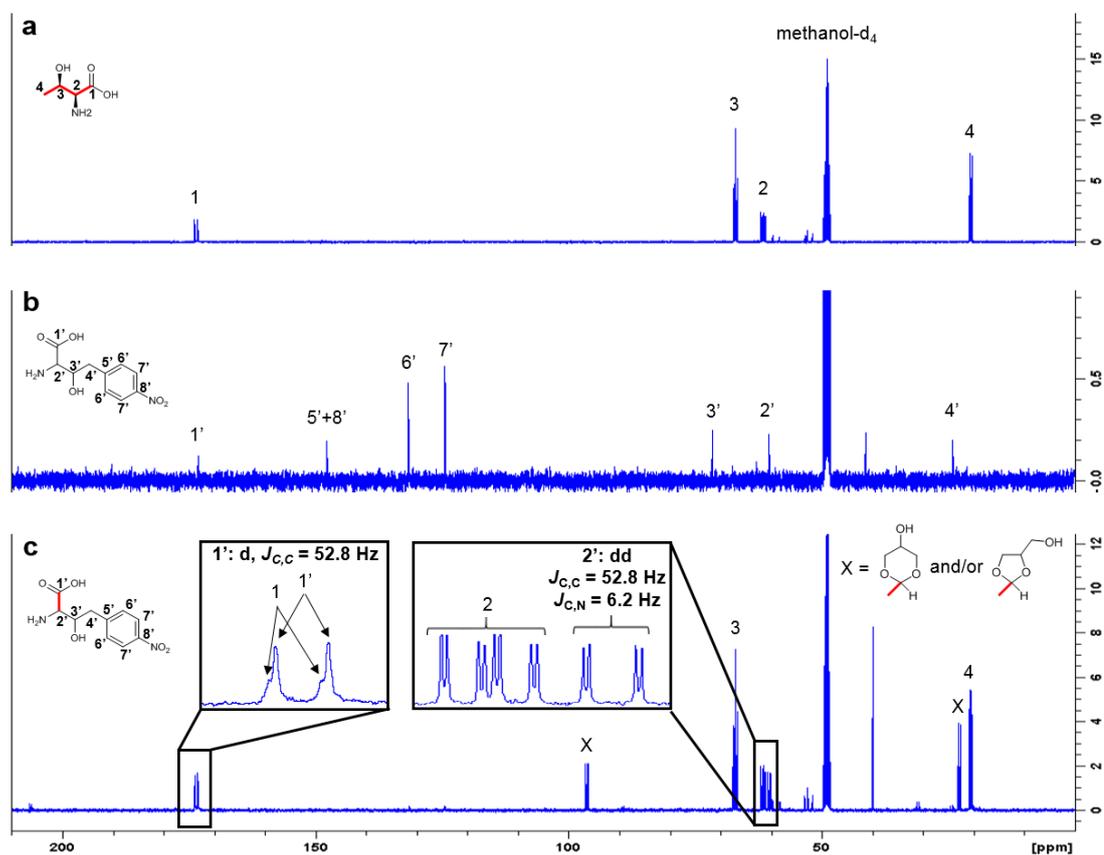


Figure 5.3. Characterisation of the biosynthesis of AHNB (2) by ObaG using ^{13}C NMR. (a) ^{13}C NMR spectrum of $[U-^{13}\text{C}_4, ^{15}\text{N}]$ L-threonine in methanol- d_4 . (b) ^{13}C NMR spectrum of the synthetic reference **2**. (c) ^{13}C NMR spectrum of enzymatically synthesised **2**. Signals marked with **X** correspond to the two acetal forms of acetaldehyde and glycerol, being formed after release of acetaldehyde, and these are both illustrated. This Figure is courtesy of Dr. Daniel Heine in the Wilkinson group.

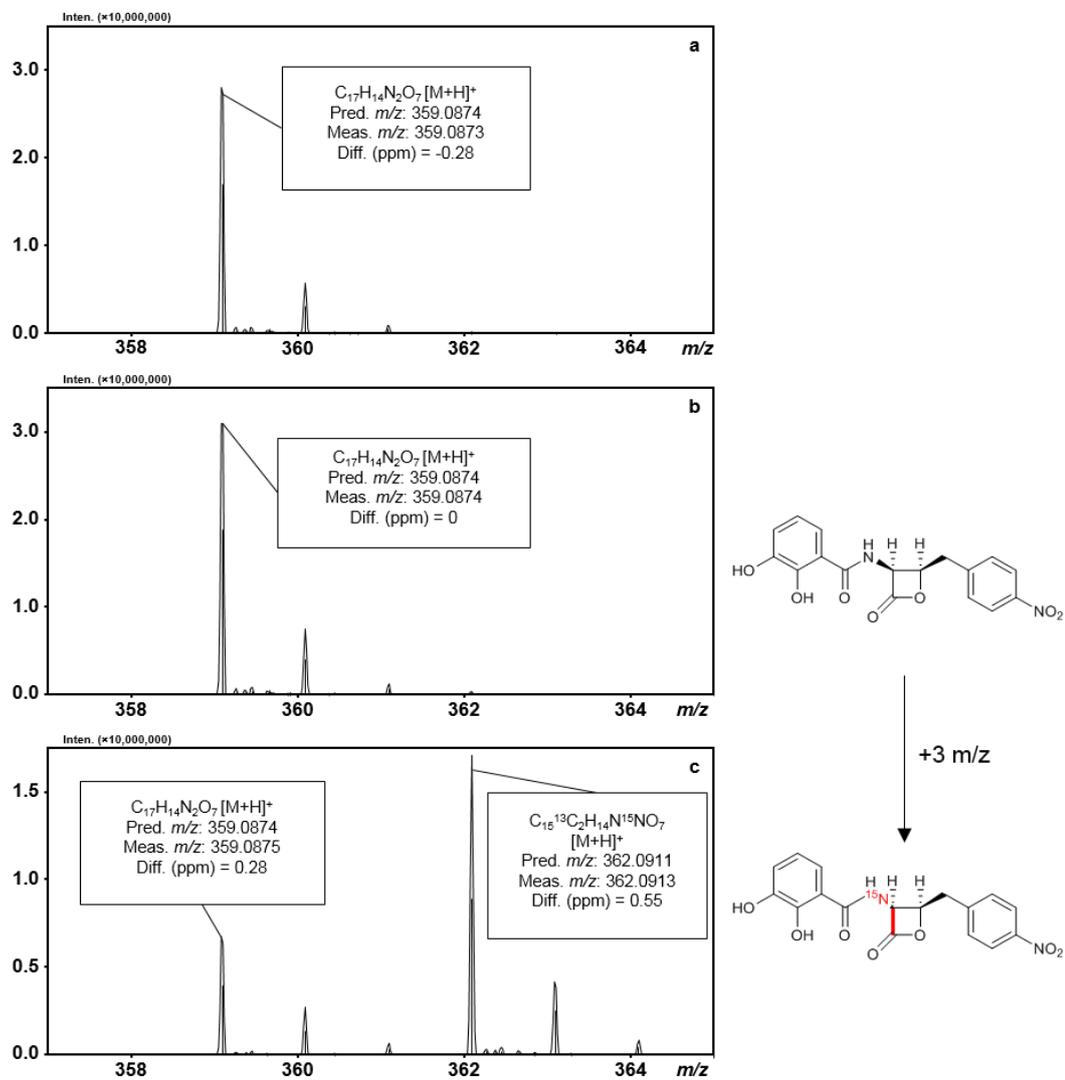


Figure 5.4. LCMS profiles for in vivo $[U-^{13}C_4, ^{15}N]$ L-threonine feeding experiments. WT production cultures fed (a) ddH_2O (negative control); (b) 2 mM unlabelled L-threonine; and (c) 2 mM $[U-^{13}C_4, ^{15}N]$ L-threonine.

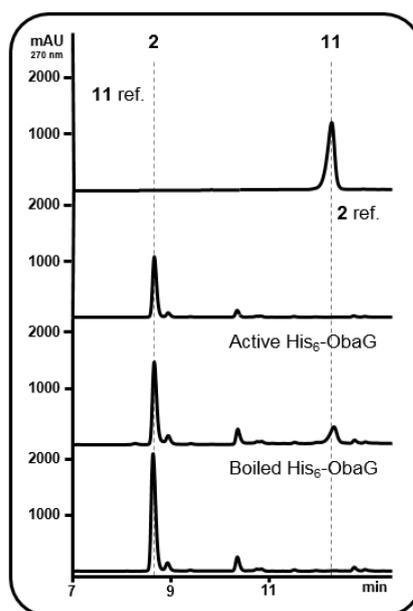


Figure 5.5. HPLC profiles for the *His₆-ObaG* reverse transaldolase assay with AHNB (2) and acetaldehyde.

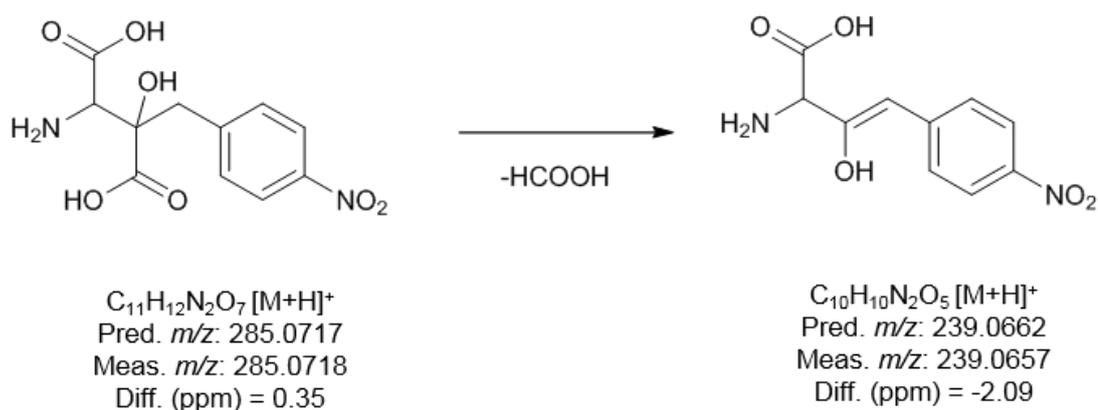


Figure 5.6. Proposed products of the *His₆ObaG* transaldolase reaction using 4-NPP (8) and *L*-threonine as substrates. A fragment for a deformedylated form of the product was also detected by MS/MS and the proposed molecule structure is also illustrated.

results (**8** only and **11** only detected respectively), but when ObaH alone was boiled, in addition to a peak matching the **8** reference, an additional peak (*) was detected (Figure 5.2d). HR-LCMS of this peak gave an accurate mass of $m/z = 285.0718$ $[M+H]^+$, consistent with it belonging to the ObaG transaldolation product of **8** (Figure 5.6), indicating that ObaG is able to tolerate it as a substrate, albeit with very low affinity.

5.5 Characterising the PLP-dependency of ObaG

Given that examples of L-TTAs are rare in the literature, experiments were performed to further biochemically characterise ObaG by confirming its identity as a PLP-dependent enzyme. An alignment was generated to compare amino acid sequences of authentic and putative SHMTs (Appendix 1 - Supplementary Figure 1). Despite initially being identified as a putative SHMT, ObaG shares only 25% sequence similarity to *bona fide* bacterial SHMTs, but does however comprise residues critical for PLP-dependent activity, most importantly the lysine residue necessary for internal aldimine formation (Schneider *et al.*, 2000). The ObaG amino acid sequence was submitted to Phyre2 for structural prediction, and the top structural match was to mouse SHMT (PDB: 1EJI). This is annotated as a 4-aminobutyrate aminotransferase-like family enzyme with a PLP-dependent transferase-like fold, and matched with 100% confidence over 92% of the ObaG query. Based on the predicted structure, ObaG comprises the large and small domains characteristic of fold-type I PLP-dependent enzymes (Schneider *et al.*, 2000). The large domain consists of a characteristic seven-stranded β sheet and the small domain folds into a 3-stranded β sheet covered with helices on one side (Figure 5.7a). PLP would be covalently attached via its phosphate group to the ϵ -amino group of the active site lysine in the large domain, located at the N-terminus of a short helix following a β strand, and would present into the active site pocket (Figure 5.7b and c). Phyre2 investigational analyses revealed that the highest degree of structural conservation between ObaG and mouse SHMT is located around the active site of both proteins (Figure 5.7d and e), which potentially reflects the structural stringency required to accommodate PLP-mediated biochemistry and would explain why the highest residue mutational sensitivities are observed in this region (Figure 5.8f).

To confirm that ObaG does indeed utilise PLP as a cofactor experimentally, two spectrophotometric assays (Materials and methods sections 2.7.4 and 2.7.5) were performed using His₆-ObaG samples in which excess PLP was removed. His₆-ObaG displays a UV/Vis spectrum characteristic of PLP-dependent proteins, comprising

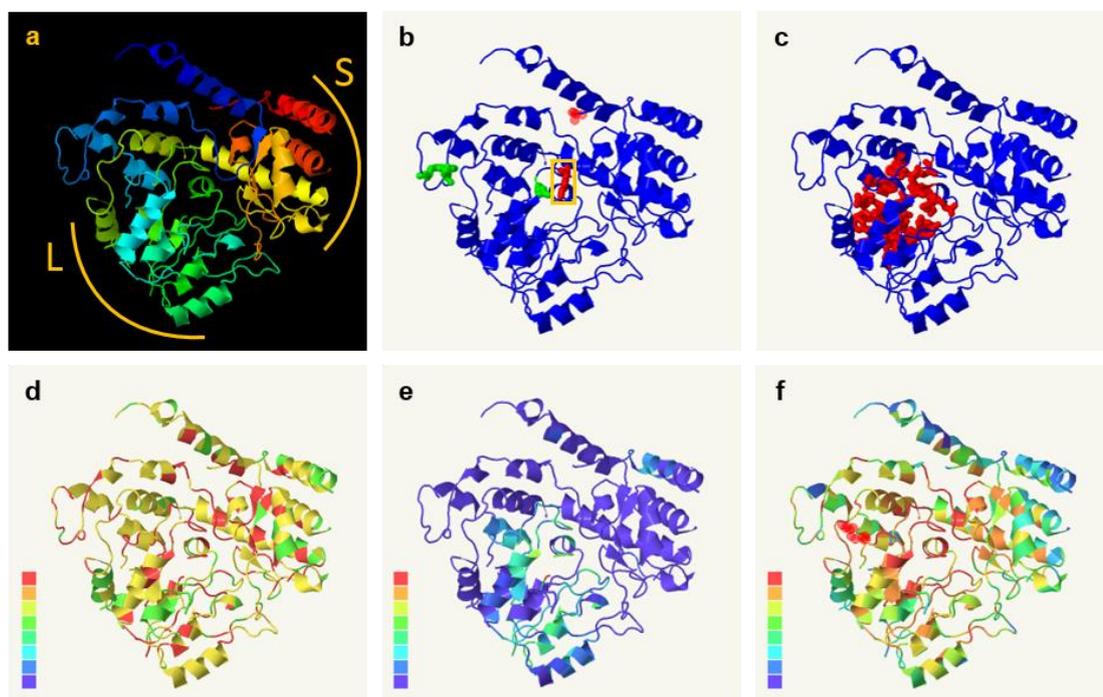


Figure 5.7. Phyre2 investigator analysis results for mouse SHMT (PDB: 1EJI).

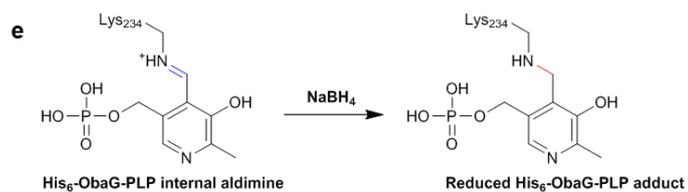
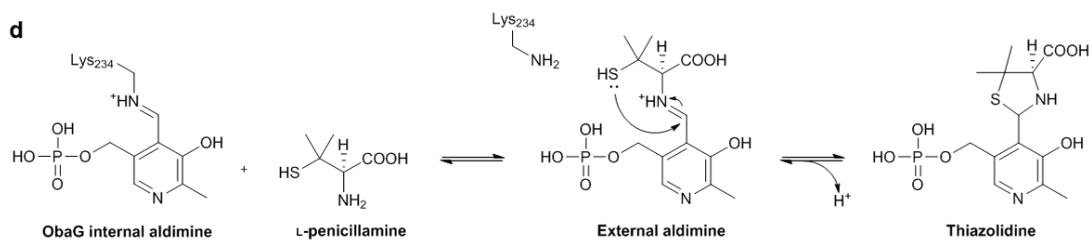
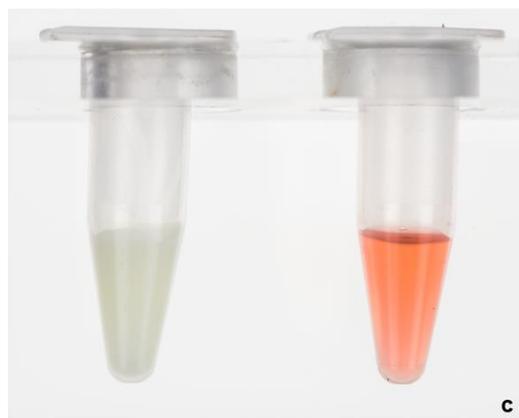
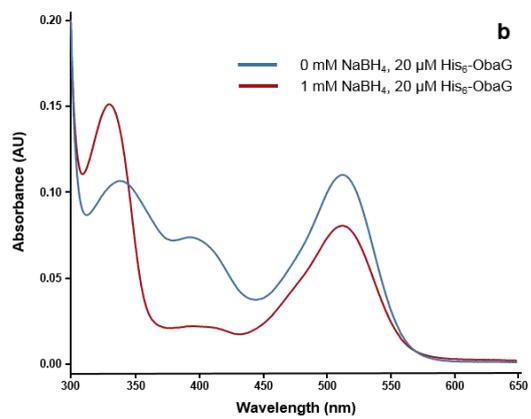
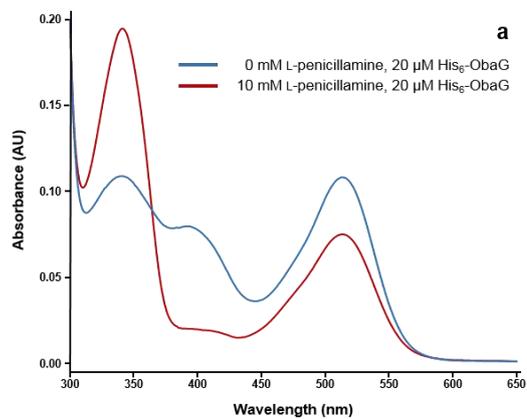
(a) 1EJI model structure, rainbow colour = N→C terminus, L indicates large domain, S indicates small domain. (b) Catalytic sites based on data from the Catalytic Site Atlas. The conserved catalytic lysine is boxed in yellow. Red indicates identical residues between query and template, green indicates different residues at this position. (c) Pocket detection as detected by fpocket2 (red indicates largest pocket). (d) Alignment confidence. The scale represents good (red) to bad (purple) confidence levels (e) Conservation based on Jensen-Shannon divergence (Capra and Singh, 2007). The scale represents high relative conservation (red) to low relative conservation (purple) (f) Mutation sensitivity as predicted by the SuSPect tool. The scale represents high mutational sensitivity (red) to low mutational sensitivity (purple).

absorption maxima at 340 and 390 nm (blue traces - Figures 5.8a and b), corresponding to the equilibrium established between enolimine and ketoenamine forms of PLP respectively. The enzyme is unusual in that it also exhibits a distinct salmon pink colour in solution (Figure 5.8c), which creates an additional absorbance maximum at 512 nm (Figures 5.8a and b). Treatment of His₆-ObaG with 10 mM L-penicillamine resulted in the anticipated formation of a thiazolidine adduct (Figures 5.9a and d), determined by UV absorbance at 340 nm, with concomitant loss of the ketoenamine peak at 390 nm. This is consistent with previous observations made for serine palmitoyltransferase (Lowther *et al.*, 2012).

In a further experiment, His₆-ObaG was incubated with 1 mM NaBH₄, resulting in the formation of a peak at 330 nm consistent with the formation of the reduced His₆-ObaG-PLP aldimine adduct (Chen and Frey, 2001 - Figures 5.8b and e), which occurs with concomitant loss of the ketoenamine peak. Finally, boiling His₆-ObaG sample in which excess PLP had been removed led to the loss of the characteristic pink colour of this enzyme, and led to the formation of a light yellow/green solution containing precipitated protein (Figure 5.8c). This likely represents the release of enzyme-bound PLP which is a yellow-green colour in solution. In summary, all results are consistent with His₆-ObaG having bound its PLP cofactor, prior to its chemical/physical removal.

5.6 Exploring the potential of His₆-ObaG for synthetic applications

With the ObaG-catalysed reaction satisfactorily characterised, preliminary kinetic data were collected to assess its potential use as a synthetic biocatalyst for the production of β -OH- α -AAs. First, a time course was determined using standard transaldolase assay conditions and showed that equilibrium between L-threonine and **11** is achieved within ~2 h following reaction initiation (Figure 5.9a). Based on these data, the reaction yields 55-60% product and as mentioned previously this is as a single stereoisomer (*L-threo*). Unfortunately, the UV/Vis spectrum of **11** overlaps with that of NADH, precluding the use of a continuous coupled assay using acetaldehyde dehydrogenase (ADH) to collect kinetic data for His₆-ObaG. ADH oxidises the acetaldehyde by-product, reducing its β -NAD⁺ cofactor to NADH in the process, and NADH has a characteristic absorbance that can be measured spectrophotometrically (Barnard-Britson *et al.*, 2012). Discontinuous methods measuring product formation with respect to varying **11** were also confounded by its inherent reactivity. The best results were finally achieved by varying L-threonine concentration using a discontinuous HPLC-based approach (Materials and methods section 2.7.2). Single-substrate kinetic analysis His₆-ObaG revealed typical Michaelis-Menten kinetics,



yielding kinetic constants of $K_m = 40.2 \pm 3.8$ mM and $k_{cat} = 62.9 \pm 1.9$ min⁻¹ (Figure 5.9b).

Preliminary experiments were also performed to see whether His₆-ObaG would accept alternative substrates. Given that the enzyme appears to be specific for L-threonine, I used benzaldehyde and phenylacetaldehyde as surrogate aromatic aldehydes to **11**, as the closely related L-TAs have been shown to accept a wide range of aliphatic and aromatic aldehyde substrates (Steinreiber *et al.*, 2007a and b). Incubation with phenylacetaldehyde led to a single product in very respectable yield (~45% based on product depletion) whose LCMS profile was consistent with the expected product (2*S*)-amino-(3*R*)-hydroxy-4-phenylbutanoate (Figure 5.10a). The reaction with benzaldehyde was less efficient leading to two products in a 1:2 ratio and poor overall yield (<20%) (Figure 5.10b). The LCMS profile of these was consistent with the production of both L-*threo* and L-*erythro*-phenylserine, respectively. An authentic standard of the DL-*threo* diastereoisomer was used for comparison.

5.7 Discussion

Mutagenetic analysis could not elucidate the roles of ObaG and ObaH in the biosynthesis of **2**, which necessitated the use of biochemical approaches to characterise their respective functions. Biochemical assays performed with both enzymes revealed that the entirely ThDP-dependent route to its production predicted by Herbert and Knaggs (1990) was incorrect. ObaH failed to produce an acyloin intermediate when supplied with **8** and glyoxylate, instead acting solely as a decarboxylase to generate **11** as an intermediate to **2**. Furthermore, phenylpyruvate was not accepted as an alternate substrate, supporting the hypothesis in Chapter 4 that ObaC-catalysed *N*-oxygenation occurs prior to ObaH activity, likely following generation of **7** by ObaF (as illustrated in Figure 4.4).

The missing biosynthetic link was shown to be ObaG, a PLP-dependent enzyme that produces **2** as a single isomer (L-*threo*) at very respectable yields (55-60%) using **11** and L-threonine as substrates. These results were reinforced by NMR analysis following biochemical assays using [U-¹³C₄,¹⁵H₁]L-threonine, which showed that ¹³C₂¹⁵N₁ is incorporated at the C-1 and C-2 positions in **2**. The two remaining ¹³C atoms are incorporated into acetaldehyde, a by-product of the ObaG-catalysed reaction. An alignment with *bona fide* and putative SHMTs, in addition to two spectrophotometric assays, confirmed that ObaG utilises PLP as a cofactor.

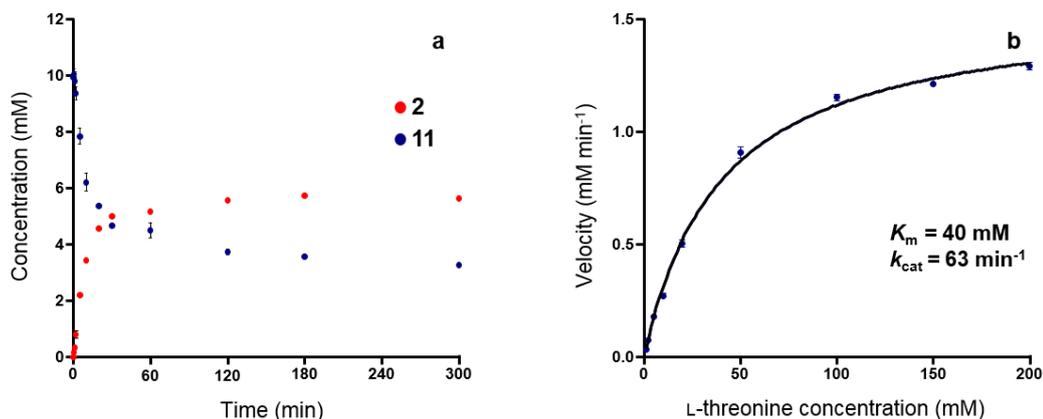


Figure 5.9. Time course and kinetic analysis of His₆-ObaG transaldolase activity. (a) Time course of the discontinuous L-TTA activity assay monitoring depletion of substrate 11 and generation of product 2. (b) Single-substrate kinetic analysis of the L-TTA reaction with 10 mM 11 and variable L-threonine (1-200 mM). All data are mean values of three (a) or five (b) independent experiments and error bars represent the s.e.m.

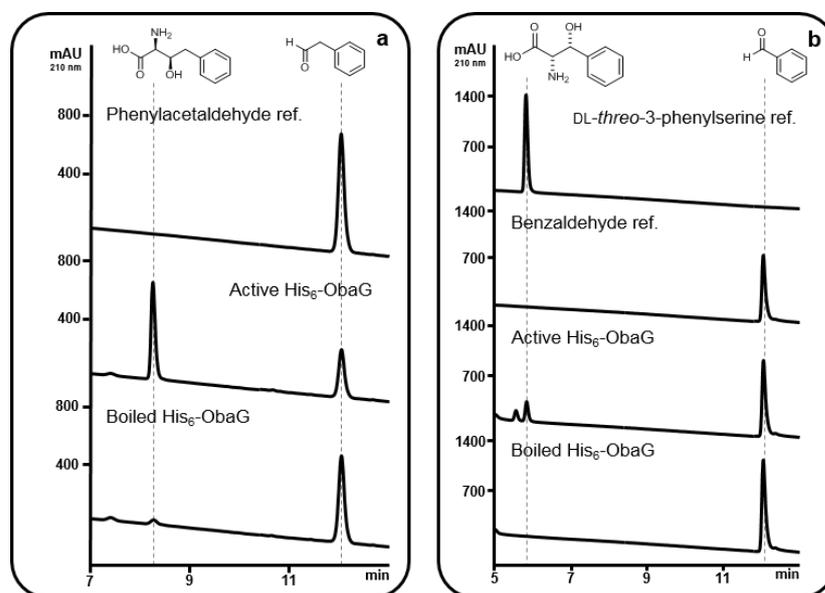


Figure 5.10. HPLC profiles for His₆-ObaG transaldolase assays with L-threonine and alternative aldehyde substrates. (a) His₆-ObaG transaldolase activity assay with phenylacetaldehyde and L-threonine as substrates. (b) His₆-ObaG transaldolase activity assay with benzaldehyde and L-threonine as substrates. Structures refer to the substrate and product references for each reaction. Reference material for (2S)-amino-(3R)-hydroxy-4-phenylbutanoate was not obtained.

Combined, these results are consistent with ObaG acting as an L-TTA which, like L-TA, initially catalyses the retro-aldol cleavage of L-threonine to generate PLP-bound glycine and acetaldehyde, the latter being released as a by-product. In L-TAs, glycine is also then released by hydrolysis, however in L-TTAs, a subsequent C-C bond forming reaction is catalysed: the glycine quinonoid intermediate undergoes an attack from an aldehyde co-substrate to generate a β -OH- α -AA product which is released to regenerate the holo-enzyme (Figure 5.11). The exact temporal nature of the ObaG-catalysed reaction remains to be elucidated, particularly with regards to when the aldehyde substrate **11** is bound for the second step of the reaction. Future work will be to obtain crystal structures for ObaG and various enzyme-substrate complexes to determine the exact mechanism of reaction, and identify critical catalytic residues that govern aldehyde substrate binding and specificity.

Herbert and Knaggs likely did not consider a PLP-dependent route for the biosynthesis of **2** because at the time of their work, no L-TTAs had been characterised, and so there was no precedent for such a mechanism. Their feeding experiments with [2-¹³C]-, [1-¹³C]- and [2-²H₂]glycine, which showed excellent specific incorporation into **1** (Herbert and Knaggs, 1990; 1992b), can be explained by the activity of SHMTs and L-TAs. Both can catalyse aldol-type reactions between glycine and acetaldehyde that would ultimately introduce stable isotope-labelled atoms into L-threonine, the substrate of ObaG. Unfortunately, the observation that glycine is readily converted to glyoxylate in *P. fluorescens* ATCC 39502, confirmed by NMR (Herbert and Knaggs, 1990), may have caused them not to consider an alternate amino acid substrate further.

L-TTAs are relatively rare and only recently have two other members of this family been described and biochemically characterised in the literature (Figure 5.12). The first to be discovered was fluorothreonine transaldolase (FTase) from *S. cattleya* (Murphy *et al.*, 2001; Schaffrath *et al.*, 2002; Deng *et al.*, 2008) which produces the antibiotic 4-fluorothreonine from fluoroacetaldehyde and L-threonine via an identical mechanism to that proposed for ObaG (Figure 5.12). The second example are members of the LipK family of L-TTAs (Barnard-Britson *et al.*, 2012), which use L-threonine and uridine-5'-aldehyde as co-substrates to produce the unusual β -OH- α -AA 5'-glycyluridine (GlyU) and acetaldehyde. Homologues of this enzyme function in a number of different biosynthetic pathways that encode lipopeptidyl nucleoside NP inhibitors of translocase I. These all incorporate GlyU as a building block, and include the caprazamycins (Kaysser *et al.*, 2009), liposidomycins (Kaysser *et al.*, 2010),

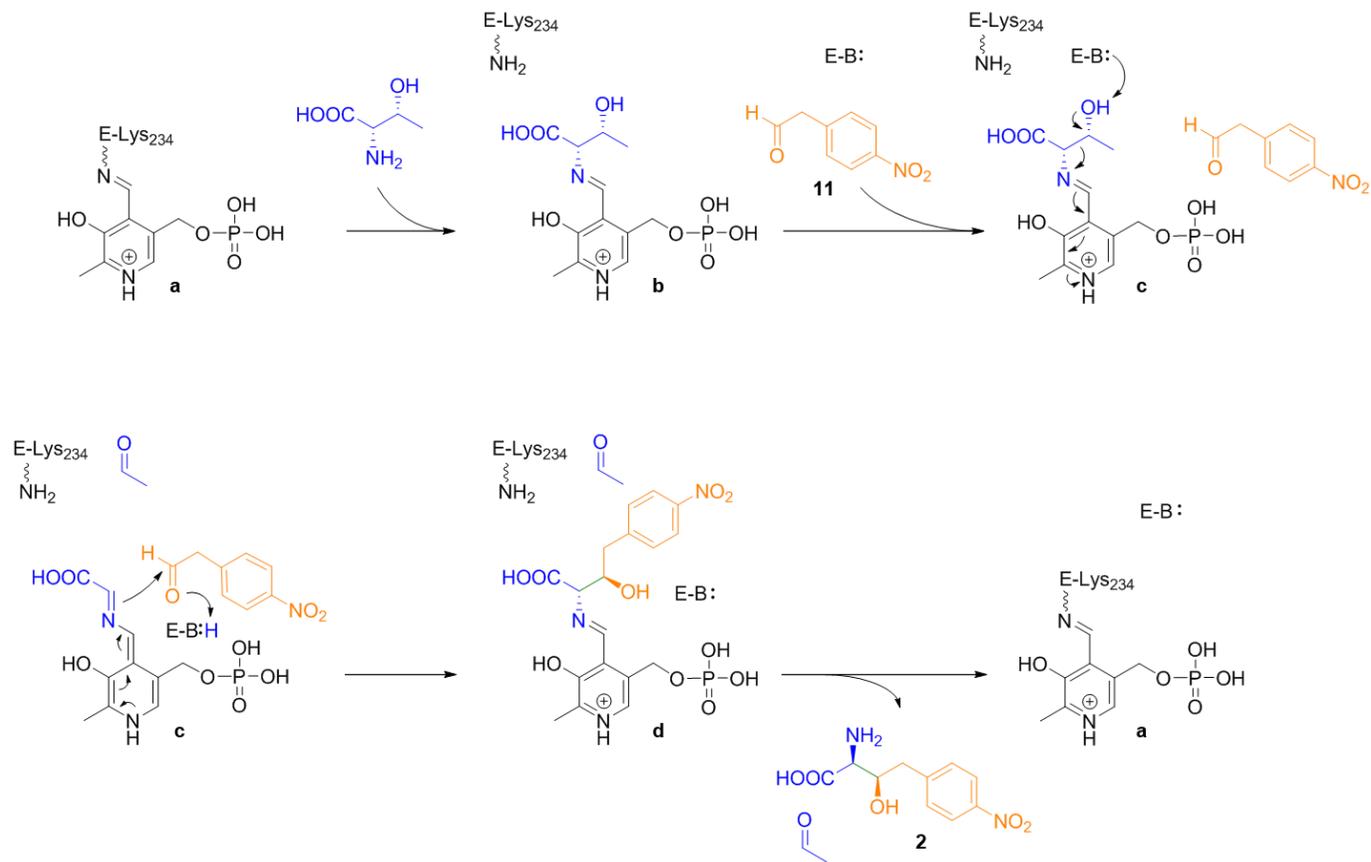


Figure 5.11. Proposed biochemical mechanism of ObaG L-TTA activity. Holo-ObaG (a) binds L-threonine (blue) to generate the external aldimine (b), and 11 (orange) to form a ternary complex. The bound L-threonine β-hydroxyl group is deprotonated by an enzyme base (c) before nucleophilic attack on the re face in an aldol-type reaction to generate the intermediate (d). 2 and acetaldehyde are then released to regenerate the holo-enzyme (a).

muraminomicin (Chi *et al.*, 2013), muraymycins (Cheng *et al.*, 2011), A-102395s, A-500359s, A-503083s, (Cai *et al.*, 2015), A-90289s (Barnard-Britson *et al.*, 2012), and sphaerimicin (Funabashi *et al.*, 2013) (Figure 5.13). Intriguingly, amino acid alignments of these representative L-TTAs with ObaG revealed relatively low shared sequence similarity, only 29% identity shared with FTase (N-terminal domain only – FTase is a didomain protein that also comprises a C-terminal phosphate binding domain) and 33% with LipK (Appendix 1 - Supplementary Figure 2). However, Phyre2 predictions indicate that they are structurally similar, all matching mouse SHMT (PDB: 1EJI) as a top hit (LipK – 100% confidence, 94% query cover; FTase – 100% confidence, 67% query cover – due to C-terminal phosphate binding domain). Convergent evolution is a common feature in the evolutionary history of PLP-dependent enzymes and has resulted in members of similar structural fold families catalysing distinct reactions, and enzymes with very different structural folds catalysing very similar reactions (Mehta and Christen, 2000; Paiardini *et al.*, 2003). It is possible that L-TTAs originated multiple times in the evolution of the fold-type I PLP-dependent enzymes, which would explain the low shared sequence similarity, or that they arose at an early stage and have greatly diverged since. Chapter 6 explores the evolutionary history of these enzymes in greater depth.

L- and D-TAs have previously been developed as biocatalysts for the generation of β -OH- α -AAs. Despite displaying relatively relaxed substrate tolerance for the aldehyde acceptor substrate, they remain limited in their use by poor stereoselectivity and low yields, even when supplied with excess glycine to shift the equilibrium in favour of the aldol reaction (Fesko *et al.*, 2016). Given that ObaG produces very respectable yields of a single optically pure stereoisomer after ~1 h under the experimental conditions used in this work, it represents a promising biosynthetic enzyme for synthetic use. However, despite exhibiting a reasonable k_{cat} , 40 mM represents a relatively high K_m , indicating that ObaG has a relatively low affinity for its L-threonine substrate. Multiple factors could have influenced this high value, including the fact that a discontinuous assay was used which will inherently introduce a greater chance of sampling errors (though relatively little standard error was observed between the five replicates at each time point), and that reaction conditions were not optimised for the enzyme from those used for analytical biochemical assays. There is also a possibility that a lower selection pressure on enzyme efficiency exists in a NP biosynthetic context (Firn and Jones, 2000), particularly for an enzyme that is not rate-limiting, compared to essential enzymes in primary metabolism that need to function with high fidelity. The validity of these results is supported by similar values of $K_m = 29$ mM and

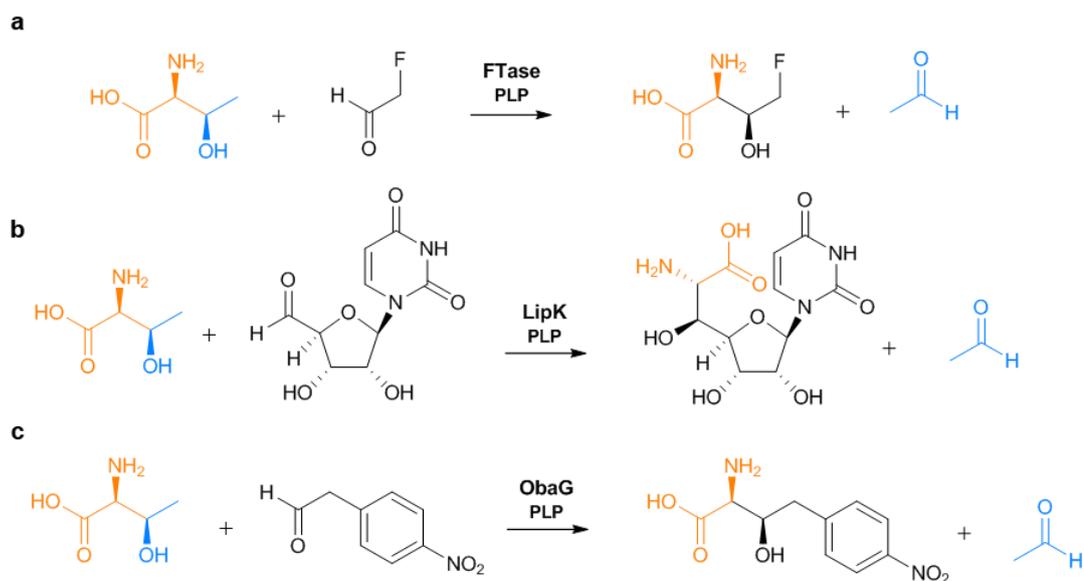


Figure 5.12. *L*-TTA-catalysed reactions characterised in the literature and this work. (a) *FTase*, (b) *LipK* and (c) *ObaG*.

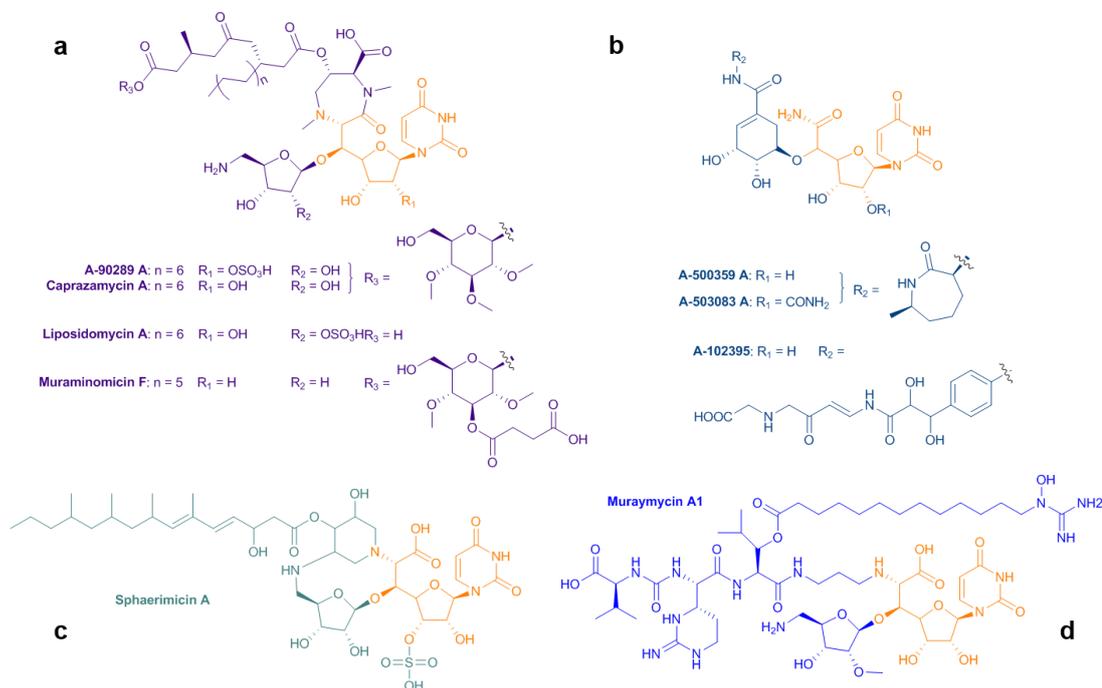


Figure 5.13. NPs comprising *GlyU* moieties derived from *LipK* homologues. (a) A-90289A, Caprazamycin A, Liposidomycin A and Muraminomycin A; (b) A-500359, A-503083 and A-102395; (c) Sphaerimycin A; and (d) Muraymycin A. The *GlyU*-derived moiety is highlighted in orange in each example.

$k_{\text{cat}} = 40 \text{ min}^{-1}$ being reported for LipK, though these were obtained using an optimised continuous assay and with respect to varying aldehyde substrate concentration (Barnard-Britson *et al.*, 2012).

ObaG was also found to display a degree of substrate promiscuity with regard to its aldehyde substrate, synthesising diastereoisomers of phenylserine from benzaldehyde, and (2*S*)-amino-(3*R*)-hydroxy-4-phenylbutanoate from phenylacetaldehyde. The latter was produced in respectable yields of ~45% product, indicating ObaG might have use in generating congeners of **2** that are substituted at the aromatic end of the molecule. Further improvements in yield might be achieved by introducing excess L-threonine and aldehyde substrates to shift equilibrium further in the direction of the biosynthesis of **2**. This reaction appears to be favoured under the experimental conditions used, based on attempts to demonstrate the reverse reaction to form **11** and L-threonine from **2** and acetaldehyde. Depletion of acetaldehyde by-product levels would also serve to drive flux through the **2**-forming reaction, and this could be achieved by introducing glycerol to react with acetaldehyde and generate acetals, as observed in [U-¹³C₄, ¹⁵H₁]L-threonine NMR analysis experiments, or by introduction of a suitable ADH to oxidise acetaldehyde to acetic acid.

As a final comment, the cause of the pink colour of His₆-ObaG remains unknown and is also present when the protein is expressed without a tag. Given that the colour is not significantly lost following L-penicillamine or NaBH₄ treatment, it appears to not be associated with PLP in the active site, and may be the cause of some other structural element or the binding of a chromophore elsewhere on the protein, although there is no evidence for this at the sequence or predicted structural levels.

Given that the SHMTs, L-TAs and L-TTAs perform analogous reactions with respect to L-amino acid cleavage to generate a PLP-bound glycine quinonoid intermediate, the next logical progression of this work was to understand how these different PLP-dependent enzymes relate to one another in an evolutionary context. Specifically, to understand how the L-TTAs diverged from SHMTs and L-TAs to utilise an aldehyde substrate, as opposed to water, to release PLP-bound glycine and form **2**. This might provide amino acid sequence-based or structural clues that allow the identification of further members of the L-TTA family. This would be highly desirable given the privileged status of β -OH- α -AAs as building blocks in both NP biosynthesis and in synthetic biology.

Chapter 6:
Phylogenetic analysis of
SHMTs, L-TAs and L-TTAs

Chapter 6: Phylogenetic analysis of SHMTs, L-TAs and L-TTAs

6.1 Introduction

PLP-dependent enzymes are thought to have originated very early in evolutionary history, before the divergence of the three biological domains of life ~1,500 million years ago. They comprise five structurally distinct fold-type families that each arose from separate evolutionary lineages (Mehta and Christen, 2000). SHMTs, L-TAs and L-TTAs belong to the fold-type I family of PLP-dependent enzymes that all act at the C_α position of their amino acid substrate, as discussed in Chapter 5. They all function as homodimers, forming two active sites at the dimer interfaces and with both monomers contributing residues to each site (Eliot and Kirsch, 2004). Although fold-type I enzymes are all structurally similar, fold-type does not dictate reaction-type, and they have greatly diversified to catalyse a broad range of different reactions. This is facilitated by the catalytic promiscuity of the PLP cofactor, which, free in solution, can catalyse multiple reactions, including transamination, racemisation and α,β -elimination, albeit at much slower rates compared to when it is acting as an enzyme cofactor. It is thus believed that ancestral enzymes could probably also catalyse a range of different reactions (Di Salvo *et al.*, 2013), and gene duplications and natural selection have driven the diversification of function among different PLP-dependent enzymes, simultaneously narrowing their substrate and reaction specificities over time.

SHMTs possess an especially broad reaction specificity. They can catalyse retro-aldol cleavage, decarboxylation, transamination and racemisation reactions, with both natural and unnatural substrates (Di Salvo *et al.*, 2013). Critically for this work, SHMTs, and the closely related L-TAs (Contestabile *et al.*, 2001), have been investigated as synthetic tools for the asymmetric preparation of unnatural β -OH- α -AAs, as both can catalyse an aldol-type reaction between glycine and an aldehyde (Figure 6.1). Whilst highly specific at the C_α of their amino acid substrates, these enzymes show relatively high aldehyde acceptor substrate promiscuity (Fesko, 2016), making them attractive targets for development as biocatalysts. Although specific for glycine, some examples do exist of L-TAs that can accept D-alanine, D-serine and D-cysteine (Fesko *et al.*, 2010), further expanding the potential to generate new chemistries. As mentioned in Chapter 1 however, these enzymes are limited in their use for the synthesis of β -OH- α -AAs by poor stereoselectivity and low yields of

the aldol reaction, and equilibrium sits in favour of the opposite retro-aldol cleavage reaction (Figure 6.1).

L-TTAs were initially identified as putative SHMT/L-TA-type enzymes, but have been shown to efficiently catalyse L-threonine β -substitutions via the sequential breaking and forming of C $_{\alpha}$ -C $_{\beta}$ bonds with aldehyde substrates, in Chapter 5 and elsewhere (Deng *et al.*, 2008; Barnard-Britson *et al.*, 2012). SHMTs, L-TAs and L-TTAs employ the same mechanism of amino acid substrate cleavage, but different selection pressures and thermodynamic arguments might explain how L-TTAs diverged to catalyse a subsequent C-C bond forming reaction (Figure 6.1). They also offer a possible explanation for the low yields and poor stereoselectivity exhibited by SHMTs and L-TAs in the reverse, aldol reaction to β -OH- α -AAs.

SHMTs are ubiquitous, essential enzymes that are necessary in folate metabolism for the generation of C $_1$ units for THF-dependent reactions, and subsequent use of those units likely drives flux through the retro-aldol cleavage reaction (Schirch *et al.*, 1985). The breakdown of a single substrate, L-serine, to two products, glycine and formaldehyde, is also entropically favourable (Figure 6.1). L-TA cleavage of L-threonine to glycine and acetaldehyde would similarly be entropically preferential, and the reactivity of the acetaldehyde by-product would likely drive the reaction equilibrium further in favour of retro-aldol cleavage through its continuous depletion. L-threonine serves as the sole source of carbon and nitrogen for growth in a wide variety of organisms and L-TAs are thought to play a key role in its catabolism, supplementing SHMT function in glycine supply if it becomes blocked (Liu *et al.*, 1998; Dückers *et al.*, 2010). The ability to effectively catalyse retro-aldol cleavage is therefore of primary importance in both cases, but identifying a selective advantage to shifting equilibrium in favour of the non-physiological aldol-type reaction is less obvious. This is reflected in the poor yields and lack of stereospecificity observed at the second chiral centre in SHMT/L-TA-catalysed aldol reactions (Fesko, 2016).

It is possible that L-TTAs evolved from the duplication of a gene encoding either an SHMT or an L-TA. Independent of the selective restrictions of primary metabolism, L-TTAs have subsequently diverged to perform an additional C-C bond forming step following retro-aldol cleavage, leading to the formation of new products in specialised metabolic contexts. Indeed, in addition to *obaG*, there are two putative SHMTs and two putative L-TAs encoded elsewhere in the ATCC 39502 genome. The C-C bond-forming reaction to generate β -OH- α -AA products has evolved as a subsequent step

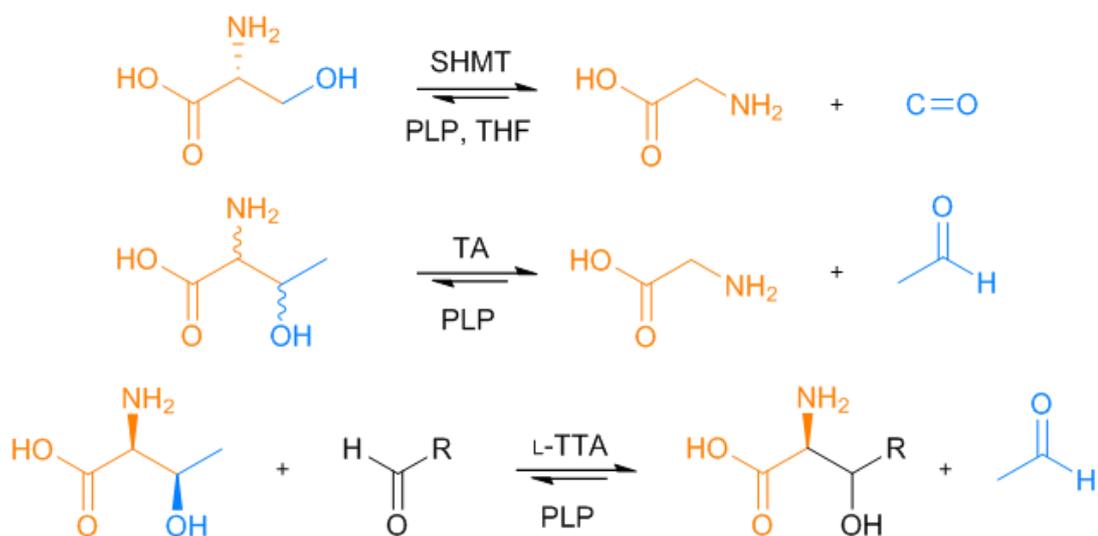


Figure 6.1. Reactions catalysed by SHMTs, L-TAs and L-TTAs. Schemes illustrating the reactions catalysed by SHMTs (top), L-TAs (middle) and L-TTAs (bottom).

to retro-aldol cleavage, rather than via evolutionary optimisation of the reverse aldol reaction catalysed by SHMTs and L-TAs (Figure 6.1), again reflecting the primary metabolic relevance of this reaction. In addition to generating β -OH- α -AA products in reasonable yields, L-TTA-catalysed reactions are stereospecific, perhaps highlighting the importance of stereochemistry for substrate recognition by downstream-acting proteins. In the case of obafluorin (**1**), changes in the stereochemistry of AHNB (**2**) could have substantial consequences for the formation and/or activity of **1**.

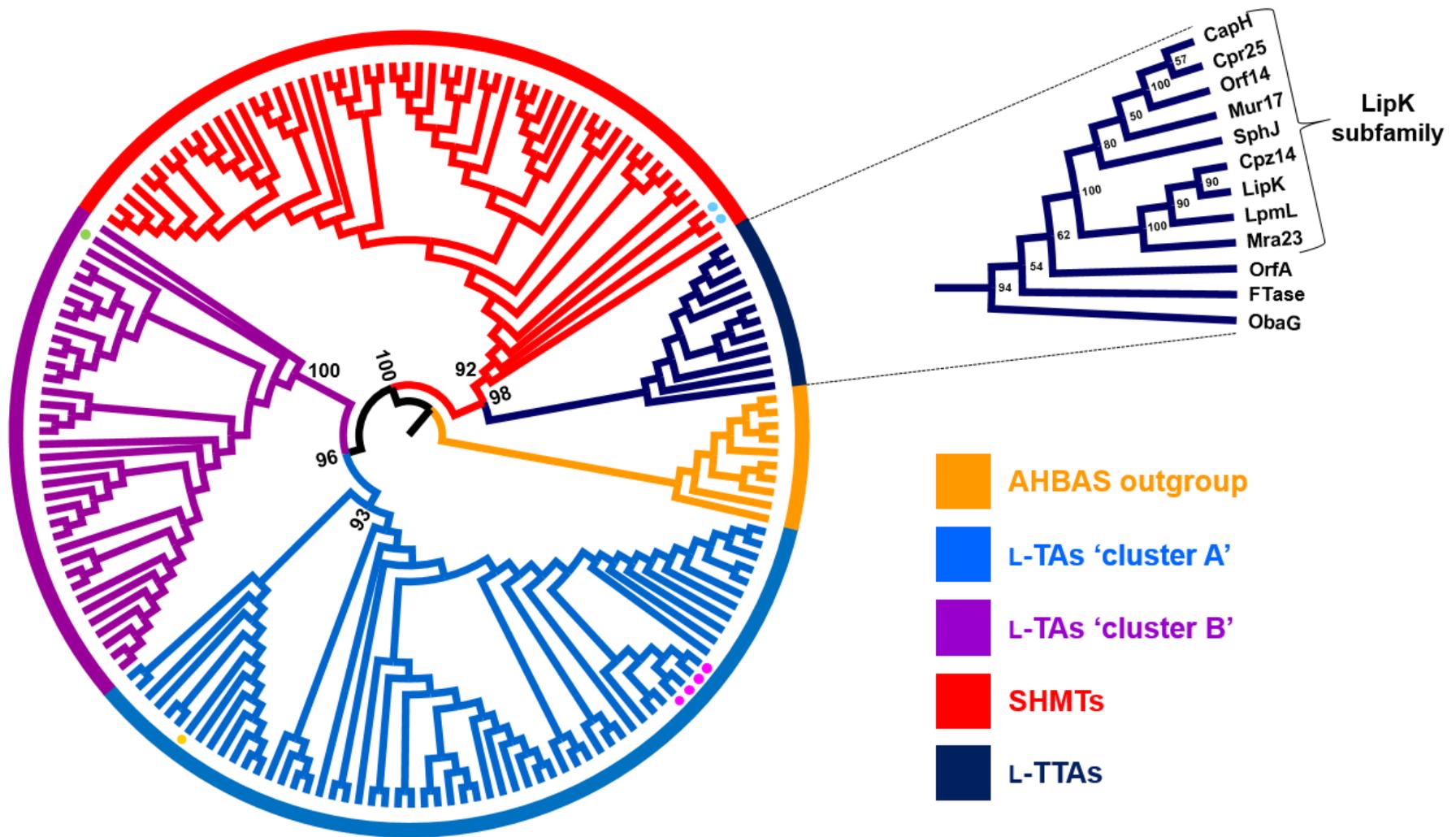
Given the close structural similarity and shared reaction mechanisms, the evolutionary relationship between SHMTs, L-TAs and L-TTAs was investigated further. D-TAs (TAs specific for D-threonine) have also been reported, and are enantio-complementary to L-TAs. However, they belong to the fold-type III family of PLP-dependent enzymes and represent an example of convergent evolution (Paiardini *et al.*, 2003). Previous phylogenetic approaches have been applied successfully to demonstrate that the L-TAs form two discrete evolutionary lineages that are phylogenetically distinct from the closely related SHMTs (Liu *et al.*, 2015). Although all PLP-enzymes share very similar mechanistic features with regards to PLP-chemistry, and the reactions catalysed by SHMTs and L-TAs are very similar, there is clearly some basis at the sequence level that allows them to be distinguished phylogenetically. If this is also the case for the L-TTAs, it creates a clear opportunity to identify further members of the family, and expand the synthetic toolkit for the preparation of unusual β -OH- α -AA building blocks. In this Chapter, the construction of a phylogeny for the fold-type I PLP-dependent enzymes described is presented. The results replicate previous findings (Liu *et al.*, 2015) and allow a hypothesis as to the likely evolutionary origins of the L-TTA family to be proposed. Furthermore, it is shown that phylogenetics may represent a powerful approach for the expansion of the L-TTA family and could aid in future genome mining efforts.

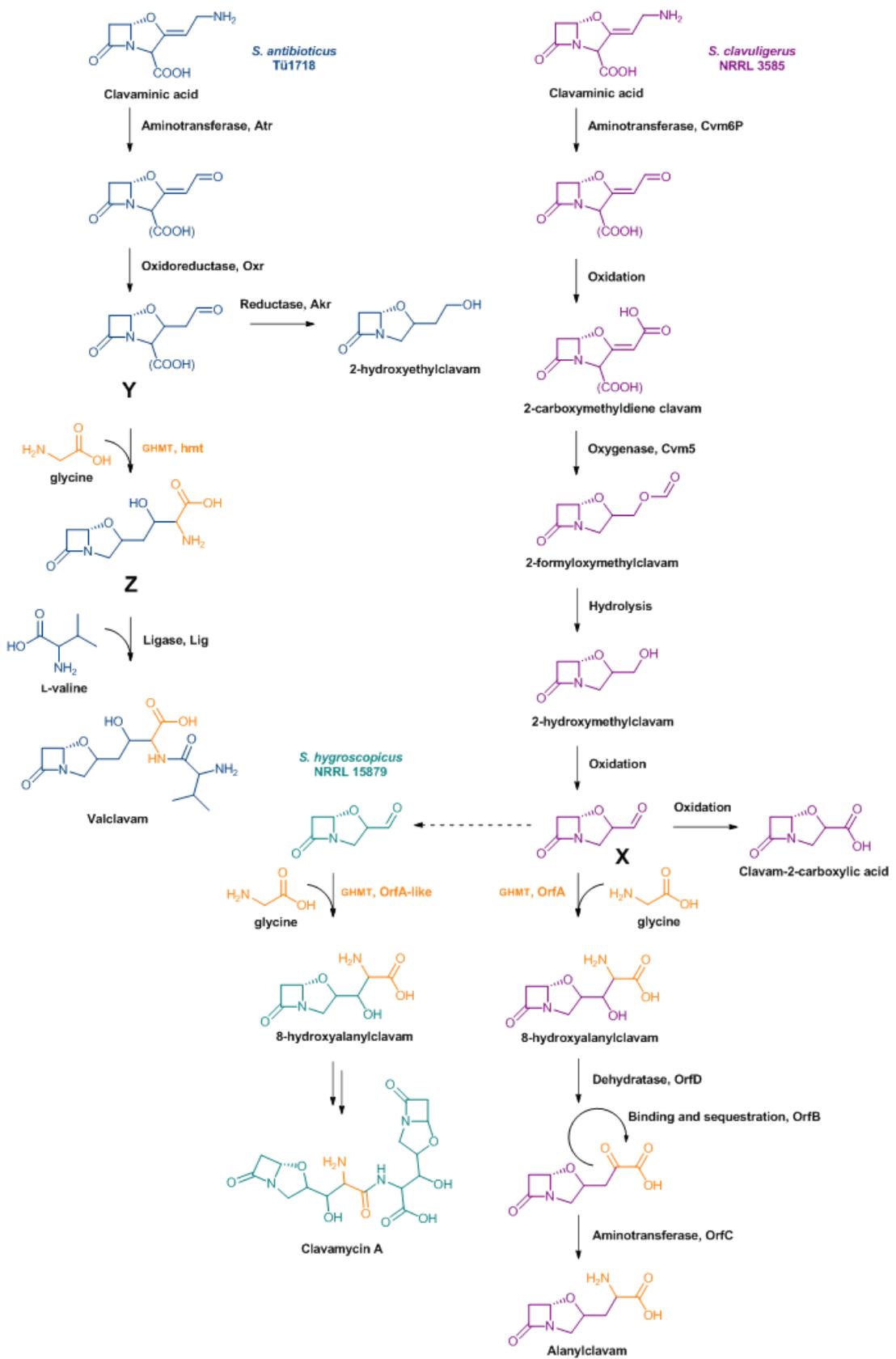
6.2 Phylogenetic analysis of SHMTs, L-TAs and L-TTAs

83 amino acid sequences of putative L-TAs and SHMTs were obtained from a previous study (Liu *et al.*, 2015) and were combined with 81 additional sequences from both enzyme families for which the crystal structure has been solved or that have been characterised/described in the literature (see Appendix 1 - Supplementary Table 2). These were collected from the GenBank database (Clark *et al.*, 2016) and the Protein Data Bank (PDB) (Berman *et al.*, 2000). A focussed search of the literature was also conducted to identify putative SHMTs, L-TAs or closely related enzymes (e.g. methylSHMTs) involved in NP biosynthesis. NP-associated enzymes are

discussed in subsequent sections. Blastp searches were performed using characterised L-TTAs as queries, and any hits sharing >30% amino acid sequence identity belonging to a published BGC were included. This cut-off was selected based on shared amino acid sequence identities between ObaG, FTase and LipK (Supplementary Figure 2) Finally, ten amino acid sequences for 3-amino-5-hydroxybenzoic synthases (AHABSs) described in the literature were obtained to function as an outgroup for phylogenetic analysis, as the AHABS family shares a recent common ancestor with SHMTs and L-TAs (Contestabile *et al.*, 2001). Selected sequences were first aligned with ClustalX2 (Larkin *et al.*, 2007) before several iterations of manual sequence trimming to remove aberrant N- and C-terminal sequence data (Materials and methods section 2.11). The sequences were ultimately aligned with T-Coffee (Notredame *et al.*, 2000), and phylogenetic tree inference was performed using maximum likelihood/rapid bootstrapping with RAxML-HPC BlackBox (8.2.8) (Stamatakis, 2014). The resulting phylogenetic tree is illustrated in Figure 6.2. Alternate trees created using slightly different alignments to ensure the final tree was consistent and robust are illustrated in Supplementary Figures 3 and 4 (Appendix 1).

Consistent with previous reports, the L-TAs formed two discrete clades which, combined with data from a sequence similarity network analysis (Liu *et al.*, 2015), indicates that these enzymes evolved independently following their divergence from a shared common ancestor. The SHMTs are monophyletic, diverging from a shared progenitor with the two L-TA clades. The L-TTAs share a common ancestor with the SHMTs and also form their own discrete evolutionary lineage. Although the evolutionary relationship between ObaG and FTase is not so easily determined when comparing different phylogenies (Appendix 1 - Supplementary Figures 3 and 4), they are clearly distinct from the L-threonine:uridine-5'-transaldolases of the LipK subfamily. Given the essential nature of SHMTs in folate metabolism, and their inherent catalytic promiscuity, it is reasonable to suggest that the earliest common ancestor of the SHMTs, L-TAs and L-TTAs likely used L-serine as a substrate. It appears that a substrate shift from L-serine to L-threonine occurred multiple times in the evolutionary history of the fold-type I PLP enzymes, giving rise to the L-TA and L-TTA clades independently. An alternative interpretation is that the original ancestor of these enzymes used L-threonine as a substrate, and a duplication event gave rise to a descendant that diverged into the SHMT lineage, with a shift in substrate specificity to L-serine, and the L-TTAs, which diverged to accept an additional aldehyde substrate. However, the essential nature of SHMTs likely precludes this latter hypothesis.





Intriguingly, among the amino acid sequences for NP-associated enzymes included in the analysis, one was consistently sorted into the L-TTA clade. This was the sequence for OrfA, identified as a putative glycine hydroxymethyltransferase (GHMT) involved in the biosynthesis of alanylclavam by *S. clavuligerus* NRRL 3585 (Zelyas *et al.*, 2008). *S. clavuligerus* is known to produce numerous clavam metabolites (Jensen, 2012), but analysis by transposon-mutagenesis showed the *orfA* mutant to be specifically deficient in alanylclavam production. OrfA was predicted to be responsible for generating β -OH- α -AA intermediate 8-hydroxyalanylclavam by fusing glycine to an unusual aldehyde substrate (**X** in Figure 6.3), in an analogous aldol-reaction to SHMTs and L-TAs. This was based on the observation that the amino acid sequence comprises the three residues necessary for glycine binding, but lacks five of the six amino acids involved in 5,10-methylene-THF binding. However, OrfA activity has never been investigated biochemically and the anticipated aldehyde substrate **X** (Figure 6.3) was never detected in Δ *orfA* cultures (Zelyas *et al.*, 2008). Based on our own observations for 4-NPA (**11**), absence of **X** could be due to the rapid breakdown of what may be an unstable aldehyde intermediate.

Further investigation of the literature revealed the presence of further OrfA-like proteins involved in the biosynthesis of two other clavam metabolites. In valclavam biosynthesis, *hmt* encodes an OrfA-homologue (68% shared amino acid sequence identity) proposed to perform a similar reaction to OrfA, but using a slightly different aldehyde substrate (includes an extra carbon in the side chain – **Y** in Figure 6.3) to produce a 9-hydroxyclavam intermediate (**Z**) (Nobary and Jensen, 2012). In clavamycin A biosynthesis, the same β -OH- α -AA moiety (derived from **X**) is present as for alanylclavam, indicating that an OrfA homologue catalyses an identical reaction before subsequent divergent modifications (Figure 6.3 - Jensen, 2012). Given that it is located in the L-TTA clade, it is possible that L-threonine may be the true substrate of OrfA and its homologues and, if so, then phylogenetics may represent a powerful tool for the expansion of the L-TTA family which currently comprises only three subfamilies. However, attempts to purify soluble OrfA using both native and codon optimised genes for expression in *E. coli* have so far been unsuccessful and this phylogenetic approach remains to be validated.

Putative L-TAs have been identified in the biosynthetic pathways to several streptomycete uridyl peptide translocase I inhibitors. The products of these pathways share structural similarities to the capuramycin-like compounds described in Chapter 5, but do not comprise moieties derived from 5'-glycyluridine (GlyU – the product of

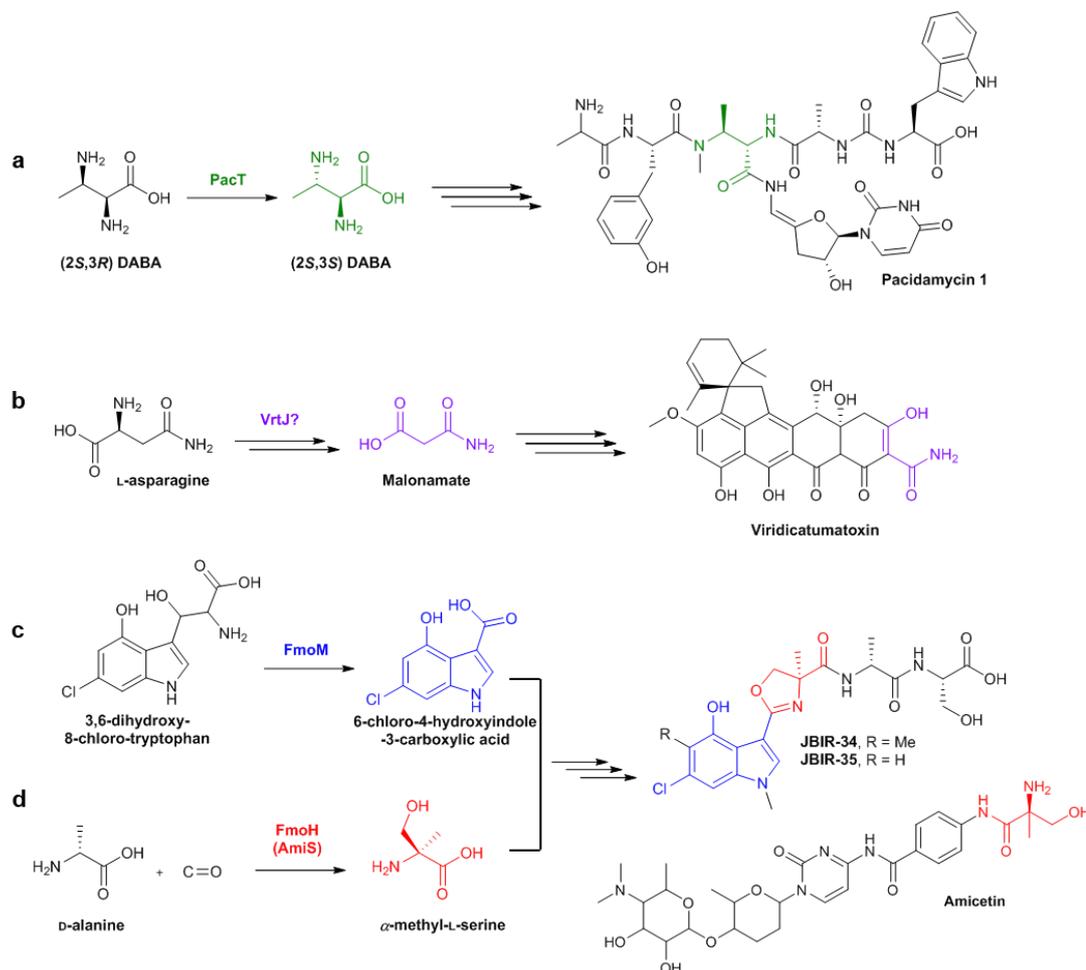


Figure 6.4. Proposed reactions catalysed by putative SHMTs/L-TAs in NP biosynthetic pathways. (a) PacT in pacidamycin biosynthesis (Zhang et al., 2010); (b) VrtJ in viridicatumatoxin biosynthesis (Chooi et al., 2010); (c) FmoM and (d) FmoH from JBIR-34 and JBIR-35 biosynthesis (Muliandi et al., 2014). AmiS, a homologue of FmoH, is responsible for the generation of α -methyl-L-serine during amicetin biosynthesis.

LipK-like L-TTAs). These L-TAs include PacT (pacidamycins – Zhang *et al.*, 2010a), 02991 (mureidomycins – Jiang *et al.*, 2015), NspJ (napsamycins – Kaysser *et al.*, 2011), and SsaT (sansanmycins – Li *et al.*, 2013), and all are hypothesised to play a role in the generation of *N*-methyl-2,3-diaminobutyric acid (DABA) in their respective pathways (Figure 6.4a). Specifically, they are proposed to epimerise DABA from the (2*S*,3*R*) configuration to (2*S*,3*S*), though this has not been biochemically validated. These four homologues consistently grouped together in all phylogenies, forming their own subclade (with 100% support) within the ‘A’ L-TA cluster (Figure 6.2), indicating that they diverged during L-TA evolution. A putative L-TA has also been identified in the biosynthesis of the hybrid PK-isoprenoid viridicatumatoxin from *Penicillium aethiopicum* Frisvad, a rare example of fungal-derived tetracycline-like compound (Chooi *et al.*, 2010). VrtJ has been identified as a decarboxylase/transaminase responsible for conversion of L-asparagine to the intermediate malonamate (Figure 6.4b), the proposed substrate of the acetoacetyl-CoA synthetase VrtB, which generates the unusual PKS starter unit malonamoyl-CoA. However, this hypothesis has not been experimentally validated and contrasts with acetate labelling results reported in the same publication. VrtJ also fell within the ‘A’ L-TA cluster (Figure 6.1), sharing a recent common ancestor with other putative ascomycete L-TAs. The *P. aethiopicum* Frisvad genome contains another copy of an L-TA-like gene (PaTA0310), which shares higher identity to Pc12g01020 (92%), the only homologue in *P. chrysogenum*, than VrtJ (54%), which implies it is the primary L-TA and supports the evolution of VrtJ as an additional copy into a specialised metabolic role (Chooi *et al.*, 2010). VrtJ could have originated either from the duplication and divergence of a primary metabolic L-TA encoded in the same genome or been acquired by HGT.

The final putative L-TA identified in a NP biosynthetic context is FmoM, thought to act canonically in catalysing the retro-aldol cleavage with subsequent oxidation of the β -OH- α -AA acid 3,6-dihydroxy-8-chloro-tryptophan during JBIR-34 and -35 biosynthesis in *Streptomyces* sp. Sp080513GE-23 (Figure 6.4c - Muliandi *et al.*, 2014). Its product, 6-chloro-4-hydroxyindole-3-carboxylic acid, is then combined with α -methyl-L-serine in a condensation reaction catalysed by the NRPS FmoA3 Cy domain before further NRPS-catalysed extensions. A second fold-type I PLP-dependent enzyme encoded in the pathway, FmoH, is implicated in the biosynthesis of α -methyl-L-serine (Figure 6.4d), and is annotated as a putative D-alanine/ α -methyl-L-serine hydroxymethyltransferase. Biochemical and stable isotope feeding experiments confirmed the role of FmoH the conversion of D-alanine and

formaldehyde to α -methyl-L-serine in an aldol-type reaction, allowing a similar function to be ascribed to a homologue in amicitin biosynthesis, AmiS (Figure 6.4d - Zhang *et al.*, 2012). FmoM diverges very early in the L-TA 'B' cluster, and FmoH and AmiS share a common ancestor that diverged early in the evolution of the SHMTs (Figure 6.2).

6.3 Discussion

Detailed phylogenetic analysis has shed light on the evolutionary origins of the L-TTA family, which appears to have diverged from a common ancestor with the SHMTs. This is perhaps not surprising given the inherent catalytic promiscuity associated with the SHMTs. The SHMTs and L-TAs in turn have evolved from a progenitor that also gave rise to the two L-TA clades that appear to have diverged independently from a shared common ancestor (Liu *et al.*, 2015). Given the essential nature of SHMT in metabolism, it is assumed that the early ancestor of these enzymes likely used L-serine as a substrate. In order to explore this hypothesis further, future work will include an ancestral state reconstruction to predict the specific amino acid substrate of the earliest common ancestor of the SHMTs, L-TAs and L-TTAs. Additionally, the construction of a species phylogeny will aid in distinguishing between duplication and divergence events, and examples in which genes encoding members of these fold-type I PLP-dependent enzyme families have been acquired by HGT. Although the exact evolutionary relationship between FTase and ObaG in the L-TTA clade could not be determined, they are clearly distinct from members of the LipK family. OrfA appears to be more closely related to the LipK family, consistently branching from a shared recent common ancestor with these enzymes (Figure 6.2 and Supplementary Figures 3 and 4).

The necessity to accommodate a second aldehyde substrate is likely to be the feature that differentiates the L-TTAs from the L-TAs and SHMTs at the amino acid sequence level. If so, it would make a phylogenetic approach ideal for the identification of novel members of this potentially useful enzyme family. OrfA, previously identified as a GHMT in alanylclavam biosynthesis (Zelyas *et al.*, 2008), was consistently placed within the L-TTA clade, indicating that it might be a genuine L-TTA. Given that the anticipated substrate of OrfA, **X**, and product, 8-hydroxyalanylaclavam, have already been proposed (Figure 6.3 - Zelyas *et al.*, 2008), this enzyme represents an ideal candidate to biochemically validate a phylogenetics-based mining approach for further L-TTAs. Unfortunately, efforts so far to purify soluble OrfA have failed with both native and codon-optimised sequences for expression in *E. coli*. The aldehyde

substrate has also proved difficult to isolate, and the biochemical assay necessary to validate a phylogenetic genome mining approach thus remains elusive.

Dr. Esther Schmitt at Novartis is currently sequencing the clavamycin producing-stain *S. hygrosopicus* NRRL 15879. She has kindly agreed to share the genome sequence to allow the identification and cloning of the OrfA-like homologue in the pathway. The valclavam-producer *S. antibioticus* Tü1718 the genome of which encodes the OrfA homologue *hmt* has also kindly been gifted to us by Prof. Wolfgang Wohleben. Future work will thus involve obtaining soluble protein for one of these OrfA homologues, isolating the necessary substrates, and proving their function biochemically (Figure 6.5). We may also be able to use soluble homologues to generate non-native clavam aldehyde products (i.e. use Hmt to generate **X** - Figure 6.3). This approach was applied to characterise the chorismate hydrolase activity of RapK and homologues FkbO₅₀₆ and FkbO₅₂₀ in the biosynthesis of the macrocyclic PKs rapamycin, FK506 and FK520 respectively (Andexer *et al.*, 2011). If all homologues prove insoluble in *E. coli*, then expression in a *Streptomyces* heterologous host is also an option, where codon usage will be less of an issue and the proteins are more likely to be correctly folded. Expression vectors for use in *Streptomyces* spp. have already been developed in our department (Takano *et al.*, 1995), and an *S. lividans* TK24 strain has been optimised with an *strR* mutation equivalent to that which has been shown to increase levels of heterologous protein production in *Bacillus* (Prof. Mervyn Bibb, personal communication).

Dunathan proposed that the specific reaction catalysed by a PLP-dependent enzyme is determined by the orientation of the cleavable bond to the plane of the PLP pyridine ring (Dunathan, 1966). However, this can only account for the reaction specificity of a PLP-dependent enzyme, and substrate specificity is determined by further active site structural features. Evolution to remodel the substrate binding pocket to accommodate an additional aldehyde substrate, whilst retaining the shape necessary for amino acid substrate cleavage, would have required significant structural alterations in the L-TTAs. Different aldehyde substrates will require distinct active site geometries and this would explain the arrangement of different subfamilies within the L-TTA clade. Whilst aligning the amino acid sequences can provide clues as to potential key residues involved in determining substrate specificity, crystal structures of one or more of the L-TTAs, bound to PLP and their substrates, would allow the critical residues and structural motifs in these enzymes to be identified. It would also allow the temporal nature of the reaction to be determined with regards to the order

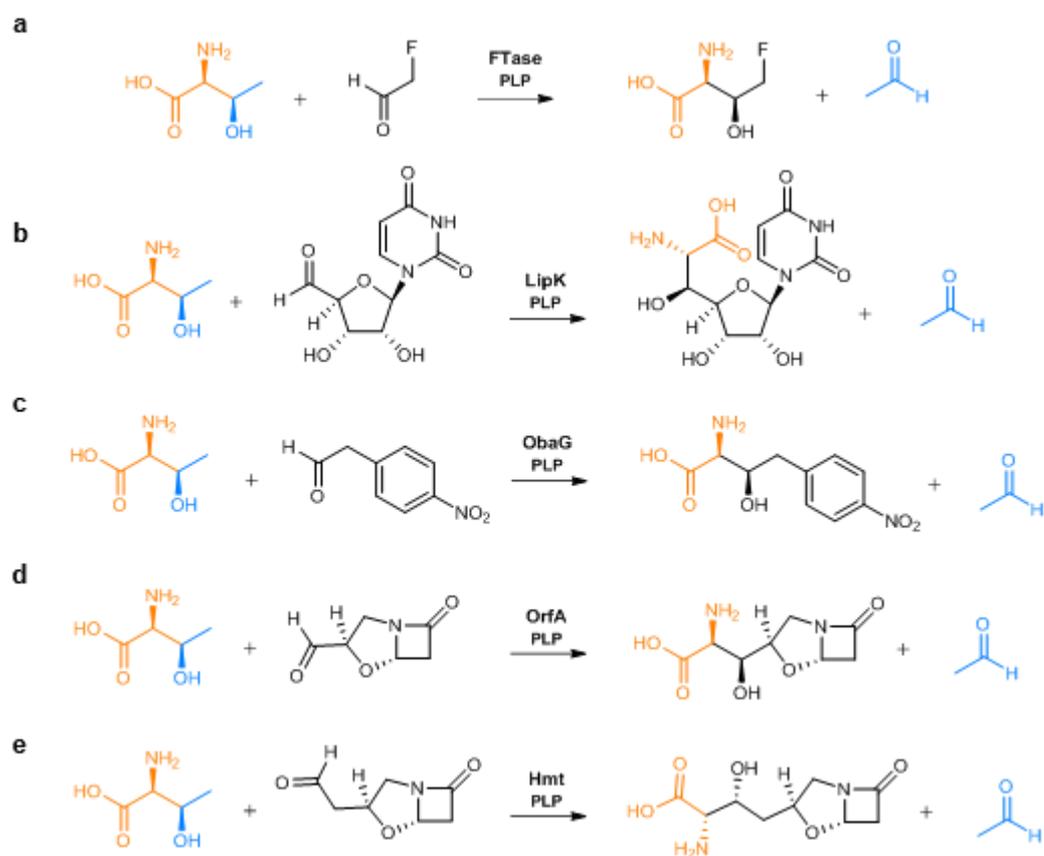


Figure 6.5. *L*-TTA-catalysed reactions characterised or proposed in the literature and this work. (a) FTase, (b) LipK, (c) ObaG, (d) OrfA and (e) Hmt.

of substrate binding. A specific search of the databases could then be performed for PLP-dependent enzymes with the specific L-TTA 'signature'. Blastp searches using LipK or OrfA queries allows homologues to be readily identified as they are the closest related matches, but below 50% identity return members of the other already known L-TTA classes. To identify L-TTAs with significantly different aldehyde substrates, and add new members to the family, an expansion of the phylogeny presented here could be performed using amino acid sequences for *all* annotated SHMTs, L-TAs and L-TTAs available in the databases as an unbiased search for novel L-TTAs. The amino acid sequences for Hmt and the OrfA homologue involved in the biosynthesis of the clavamycins will also be added to see whether they clade with the L-TTAs.

Whilst other NP biosynthesis-associated SHMTs and L-TAs were identified in this work, none of them was revealed by phylogenetics to belong to the L-TTA family of PLP-dependent enzymes. However, together with the L-TTAs, they serve to illustrate the impressive catalytic versatility of the PLP-cofactor and the broad range of functions it can bestow on enzymes. This is especially true in NP biosynthesis where there are many examples in which PLP-dependent enzymes have been acquired by BGCs and perform novel specific functions. These range from canonical transaminations, as observed for myxochelin (Silakowski *et al.*, 2000) and zeamine biosynthesis (Masschelein *et al.*, 2013), to far more unusual functions such as the PLP-dependent chain release mechanisms characterised for fumonisin biosynthesis in *Fusarium verticillioides* (Gerber *et al.*, 2009), and proposed for prodigiosin biosynthesis in *Serratia* and *Streptomyces* spp. (Garneu-Tsodikova *et al.*, 2006). Embedded PLP-dependent tailoring domains are widely distributed in modular PKSs and NRPSs (Milano *et al.*, 2013), but few, such as the aminotransferase domain of *MycA* in mycosubtilin biosynthesis (Aron *et al.*, 2005), have been characterised biochemically. Clearly, nature has realised the functional potential of PLP-dependent catalysis, and it is hoped that phylogenetics-based approaches might allow us to identify further unusual PLP-dependent biosynthetic enzymes and realise their synthetic potential. The PacT homologue subclade (Figure 6.2) is promising in this regard.

Chapter 7:

Biochemical characterisation
of obafluorin (**1**) assembly and
 β -lactone ring formation

Chapter 7: Biochemical characterisation of obafluorin (1) assembly and β -lactone ring formation

7.1 Introduction

With the roles of ObaH and ObaG in the biosynthesis of AHNB (**2**) biochemically characterised, the focus of the project shifted to investigating how **2** is subsequently incorporated into obafluorin (**1**) (Figure 4.4). Mutagenesis experiments supported the initial hypothesis that both Obal and ObaK fulfil this function, and together they comprise all the domains necessary to catalyse peptide bond formation between **2** and 2,3-DHBA (**6**), before releasing **1**. Obal represents a particularly non-canonical domain organisation that could be due to the presence of an unusual embedded MbtH-like domain. It may have facilitated the recruitment of the **6**-specific A₁ domain to the C-terminus of Obal to permit the MbtH-like domain to interact with both Obal A domains. This organisation would explain the apparent domain splitting of an EntB-like progenitor to generate discrete ArCP ObaK and isochorismatase ObaJ proteins, affording ObaK increased mobility to better interact with both the unusually located A₁ domain and the C₂ domain.

As discussed in Chapter 3, MbtH-like proteins are traditionally auxiliary proteins that are required for the activity of many, but not all NRPS A domains. Genetic knockout experiments have demonstrated their importance for the efficient production of NRPs and in some cases they have been shown to be critical for A domain functional expression (Lautru *et al.*, 2007; Wolpert *et al.*, 2007; Heemstra *et al.*, 2009), suggesting a chaperone like function. There is growing evidence that MbtH-like proteins could also be involved in allosteric regulation of A domains and so may influence catalytic properties (Herbst *et al.*, 2013; Miller *et al.*, 2016). Remarkably, MbtH-like protein paralogues can partially complement one another, and this phenomenon has also been observed *in vitro* for MbtH-like proteins from different species, even for A domains with different amino acid substrates (Wolpert *et al.*, 2007; Zhang *et al.*, 2010b; Boll *et al.*, 2011). Whilst structural data and mutagenesis experiments have identified amino acid residues critical for interaction with A domains (Herbst *et al.*, 2013; Miller *et al.*, 2016), the exact mechanism of MbtH-like protein activity remains obscure. Given that an MbtH-like domain is embedded within Obal, it is reasonable to assume that it is functionally significant for one or both A domains. If soluble and active protein can be obtained, Obal may thus represent an ideal

system for further studying the role of MbtH-like proteins and domains in NRP biosynthesis.

Beyond activation and assembly of **2** and **6**, the formation of the β -lactone ring during **1** biosynthesis also remains to be addressed. The electrophilic β -lactone moiety is a privileged structure among NPs, as it is susceptible to attack from nucleophilic amino acid residues in protein active sites (Figure 1.13), making these compounds potent and specific inhibitors of important protein targets (De Pascale *et al.*, 2011). Identifying new enzymatic routes to the formation of these strained four-membered rings is thus highly desirable, particularly when considering the closely related β -lactams, which themselves have served as a particularly rich source of marketed drugs and have opened up new areas of target-based drug discovery (Kluge and Petter, 2010).

Considerable effort has been put into identifying and characterising the biosynthetic pathways to β -lactones as a platform to developing synthetic routes to novel molecules. However, in many cases the mechanisms of ring formation remain unclear, and no enzymatic domains have been experimentally linked to cyclisation in PKS or NRPS-derived examples. The β -lactone rings in lactacystin (Dick *et al.*, 1996) and ebelactone (Wyatt *et al.*, 2013) have been shown to form nonenzymatically at pH 7 from β -hydroxythioesters, challenging the evolutionary need for enzyme-catalysed ring formation. However, **1** is readily hydrolysed to **10** under aqueous conditions, which could necessitate an enzymatic mechanism of β -lactone ring formation to ensure sufficient titres of the active metabolite are achieved. The β -lactone synthetase OleC in olefin biosynthesis remains the only characterised enzymatic route to these molecules, with further homologues identified in the ebelactone and lipstatin clusters (Christenson *et al.*, 2017), though their role in catalysing ring formation in these latter pathways has not yet been characterised experimentally.

An alternative enzymatic route to β -lactones is via NRPS C domains that lack the full catalytic triad for condensation (Cy domains), and which catalyse cyclisations to form heterocycles. Cy domains do this by catalysing intramolecular amide bond formation with cysteine, serine or threonine residues to form five-membered oxazoline and thiazoline rings. A Cy domain has been proposed to catalyse β -lactone ring formation from a serine residue in oxazolomycin biosynthesis, (Zhao *et al.*, 2010) but this reaction has not been experimentally characterised.

A putative PK cyclase is located downstream of *obaN* that belongs to the SRPBCC superfamily that includes the PYR/PYL/RCAR-like plant receptor proteins and aromatase/cyclase domain containing proteins, such as that involved in tetracenomycin biosynthesis (Caldara-Festin *et al.*, 2015). However, the mechanism of cyclisation they employ does not apply in the context of **1** biosynthesis. Given that there is no OleC homologue associated with the *oba* BGC, and the Obal C₂ domain comprises the full complement of residues necessary for catalysing condensation reactions between **2** and **6**, a TE-mediated mechanism was determined to be the most likely route to ring closure during the biosynthesis of **1**.

In NRPS assembly lines, TE-domains typically mediate product release by hydrolysis or macrocyclization, and belong to the α/β -hydrolase superfamily which contain a conserved serine-histidine-aspartate catalytic triad (Nardini and Dijkstra, 1999). In canonical NRPS biosynthesis, when the linear peptide chain reaches its full length, it is transferred from the last PCP to the hydroxyl group of the TE active site serine residue to form a peptidyl ester (Du and Lou, 2010). Depending on the nature of the TE, the peptidyl-O-TE intermediate is then attacked either by 1) an external nucleophile, typically water, which leads to a linear hydrolysed product (e.g. vancomycin – Hubbard and Walsh, 2003); or 2) by an internal nucleophile, typically an amino or hydroxyl group on the intermediate that leads to macrolactam (e.g. tyrocidine – Trauger *et al.*, 2000) or macrolactone (e.g. enterobactin – Frueh *et al.*, 2008) release respectively (Figure 7.1). For the biosynthesis of **1**, the internal Obal type I TE domain is proposed to catalyse β -lactone ring formation using the β -hydroxyl group of the **2**-derived moiety as an internal nucleophile (Illustrated in Figure 4.4).

In this Chapter, data are presented showing that soluble hexahistidine-tagged Obal can be obtained and is functional *in vitro*, as determined by A domain substrate specificity assays. These experiments validate **2** as the preferred substrate for the Obal A₂ domain, in addition to revealing varying degrees of substrate promiscuity for both A₁ and A₂ domains. Experiments prepared to characterise Obal TE-mediated β -lactone ring formation and further work to reconstitute the biosynthesis of **1** *in vitro* are also described. Unfortunately, similar experiments were published during the course of the project and those results are also summarised here. This line of biochemical experimentation was terminated at this point.

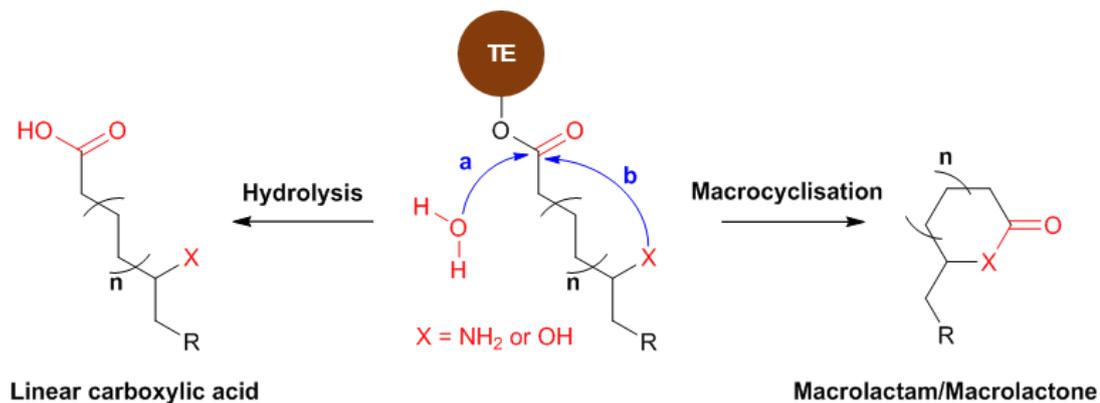


Figure 7.1. Thioesterase-mediated PK/NRP chain release mechanisms. The acyl-O-TE can either be hydrolysed by H_2O to produce a linear product (a), or macrocyclised following intramolecular nucleophilic attack to generate a macrolactam or macrolactone (b).

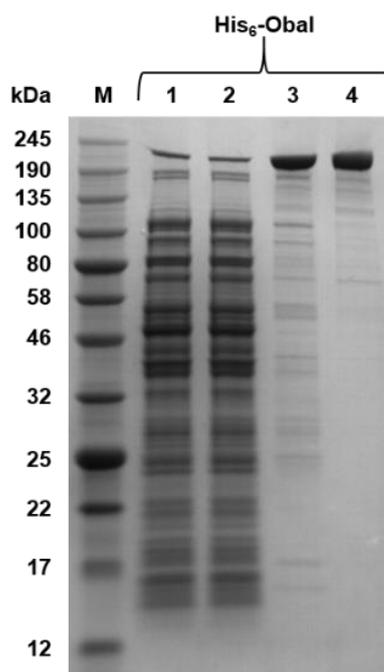


Figure 7.2. 12 % SDS PAGE gel of His₆-Obal purification steps. 1 = Cell lysate, 2 = Chitin wash eluent, 3 = Ni-affinity purified sample, 4 = Size exclusion purified sample, and M = Colour Prestained Protein Standard, Broad Range (NEB). The expected molecular weight for His₆-Obal is 211.5 kDa.

7.2 Investigating His₆-Obal A domain specificity

Traditionally, individual catalytic domains from multi-modular NRPS and PKSs are purified for biochemical characterisation experiments as it can be challenging to achieve soluble protein for entire synth(et)ases, which are often several hundred kDa in size. Recently however, the monomodular NRPSs SrfA-C from surfactin biosynthesis, EntF from enterobactin biosynthesis and AB3403 from an uncharacterised pathway in *Acineobacter baumannii*, were successfully purified and crystallised (Drake *et al.*, 2016). The entire *obal* PCS was cloned as an *NdeI-XhoI* fragment into pET28a(+) and was expressed to generate soluble *apo*-His₆-Obal (211.5 kDa, ~2.5 g/L - Figure 7.2). Access to the full Obal protein creates opportunities to study the dynamics between different domains and to reconstitute **1** assembly *in vitro*.

A domains catalyse the activation of amino or carboxylic acids in an ATP-dependent manner to form an (amino)acyl-AMP adenylate and PPi (Figure 7.2a). The local concentration of pantetheinyl thiolate is sufficient for an adjacent PCP domain to capture the (amino)acyl-AMP in the A domain active site, yielding aminoacyl-S-PCP (Fischbach and Walsh, 2006). In order to assess the A domain substrate specificity of soluble *apo*-His₆-Obal *in vitro*, a non-radioactive approach was used which employs the propensity of (amino)acyl-AMP intermediates to react with hydroxylamine to form hydroxamic acids (Materials and methods section 2.7.6 - Kadi and Challis, 2009). Hydroxamate adducts are easily measurable by spectrophotometry after addition of ferric iron (Figure 7.3b). In this experiment, **2** was tested against a panel of alternate L-amino and carboxylic acid substrates. **6** could not be used as a positive control because it chelates ferric iron directly and when tested led to saturated absorbance measurements. Benzoic acid (BA) and 3-hydroxybenzoic acid (3-HBA) were used as surrogates for **6** due to successful results in mutasynthesis experiments (described in detail in Chapter 8). Whilst this approach would not allow the independent dissection of the activities of the two Obal A domains, it did suggest that **2** is indeed the preferred substrate for the A₂ domain, giving the highest absorbance measurement at 540 nm of any tested acid substrate (Figure 7.4). This domain appears to exhibit some degree of relaxed substrate specificity as significant absorbances were also observed for L-serine and L-tryptophan. BA and 3-HBA gave significant signals, indicating that they are substrates for the A₁ domain, and the difference in absorbance levels between them reflects their relative successes in mutasynthesis experiments (Chapter 8). Based on this data, I

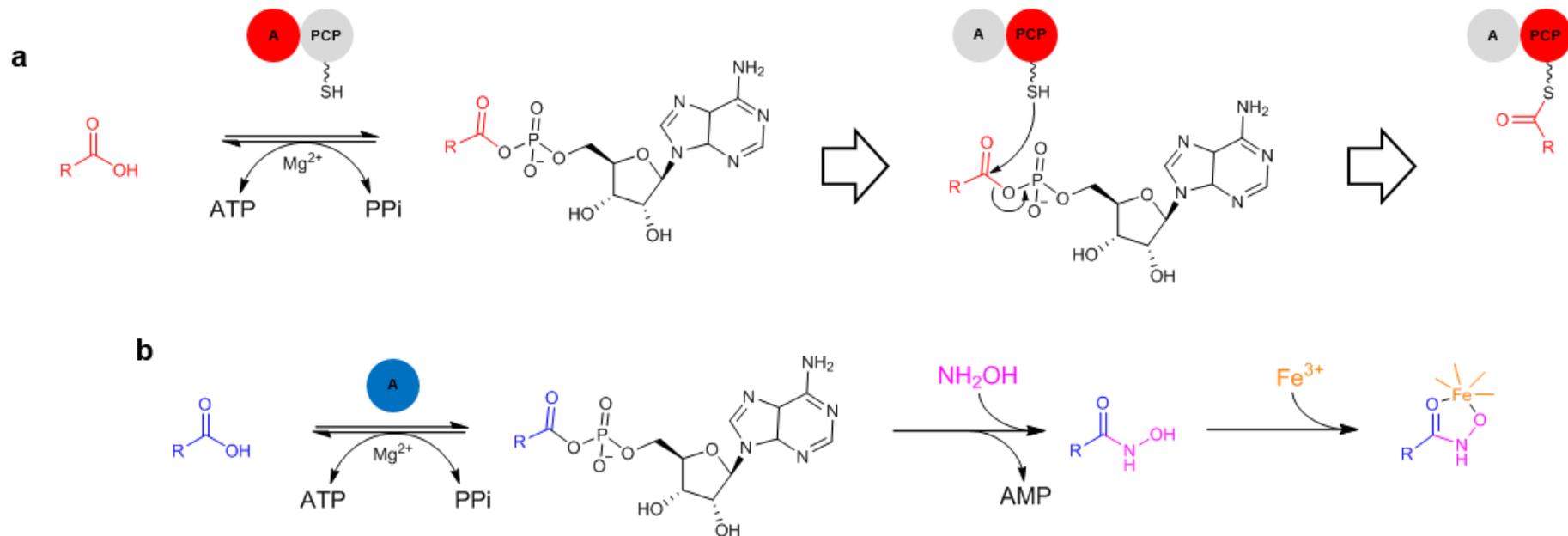


Figure 7.3. NRPS A domain-catalysed reaction mechanism and substrate specificity assay principle. (a) A domain-catalysed substrate activation. A domains first catalyse the Mg-ATP-dependent adenylation of carboxylic acid substrates. This generates an (amino)acyl-AMP intermediate that is subsequently loaded onto the cognate PCP PPant arm. Reactions illustrated are catalysed by the domain highlighted in red. **(b) Hydroxylamine-trapping assay.** Hydroxylamine intercepts aminoacyl-adenylates to form hydroxamic acids. These form complexes with ferric iron that can be detected spectrophotometrically at $\lambda_{540\text{ nm}}$ and is thus a measure of (amino)acyl-AMP formation. This will only occur when the A domain recognises and activates its cognate substrate.

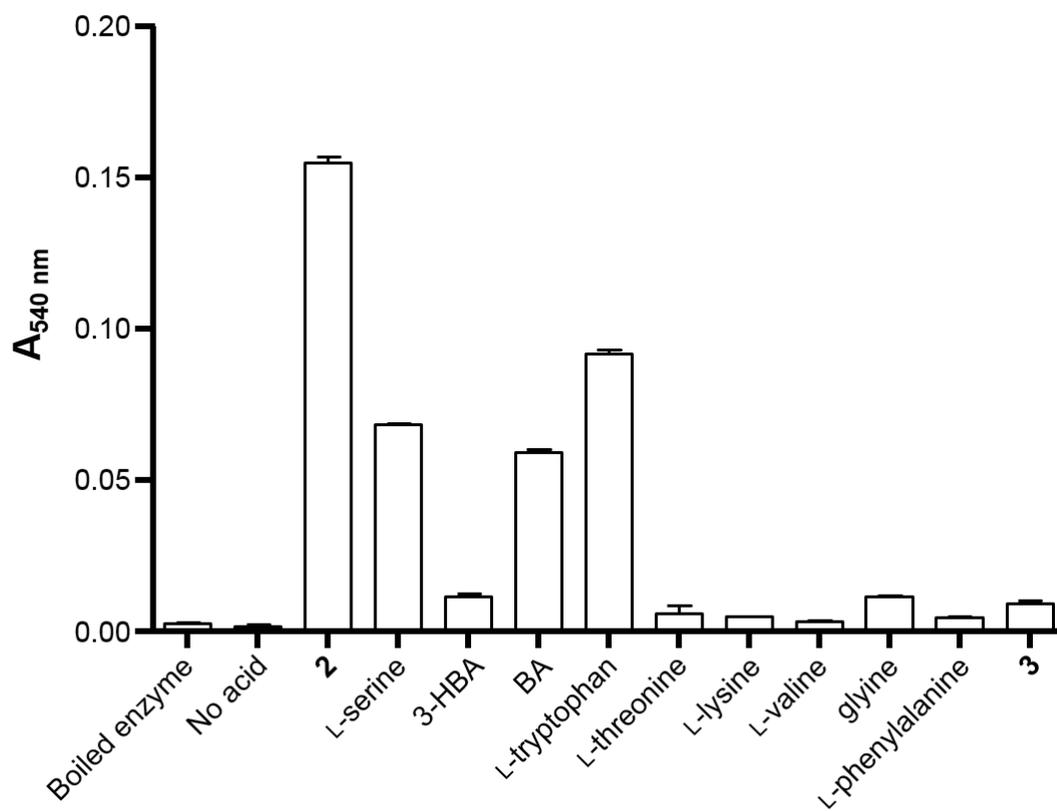


Figure 7.4. Hydroxylamine-trapping assay results. His₆-Obal adenylation domain specificity was assayed using a range of amino and carboxylic acid substrates. All data are mean values of three independent experiments and error bars represent the s.e.m.

concluded that the soluble His₆-Obal *apo*-enzyme was indeed correctly folded and functional *in vivo*, creating the possibility of studying the role of the MbtH-like domain further in the future with respect to both A domains.

7.3 Characterising His₆-Obal TE-mediated β -lactone ring formation

The ultimate objective in this line of investigation was to reconstitute the biosynthesis of **2** and the assembly of **1** *in vitro*. This would involve incubating **2** and **6** with *holo*-forms of *ObaCGHI* and *K*. To this end, *ObaH* and *ObaG* reactions were successfully coupled (Chapter 5), and *ObaC* and *ObaK* were purified as N-terminally hexahistidine tagged soluble proteins (Figure 7.5). **1** assembly would also require His₆-Obal to be phosphopantetheinylated to its *holo*-form to allow substrate tethering to PCP domains. Traditionally, this is performed using *Sfp*, the phosphopantetheinyl transferase from *Bacillus subtilis* (Quadri *et al.*, 1998b). Whilst sourcing the *sfp* gene, experiments were prepared to confirm and characterise TE activity using *apo*-His₆-Obal.

A similar approach to that used to characterise the cyclisation activity of the TE domain from *TycC* in the tyrocidine biosynthetic pathway was selected (Trauger *et al.*, 2000). Trauger *et al.* synthesised a synthetic peptide *N*-acetylcysteamine (NAC) thioester (peptide-SNAC), which was substituted for the terminal decapeptide-S-PCP to study TE domain activity (Trauger *et al.*, 2000). NAC is structurally identical to the terminal portion of PPant and is thus an excellent mimic of the natural TE substrate (Aggarwal *et al.*, 1995). Dr. Daniel Heine kindly converted **1** into the desired dipeptide-SNAC (Figure 7.6), which was confirmed by HR-LCMS and NMR, to use as a substrate for *apo*-Obal. As a negative control for the experiment, a copy of His₆-Obal in which the TE active site serine is mutated to an unreactive residue was required. To identify the correct residue, an alignment was performed of the Obal TE domain against several NRPS TE domains described and/or characterised in the literature (Appendix 1 – Supplementary Figure 5). Intriguingly, the Obal TE active site possesses a cysteine rather than a serine residue, presumably tethering the **1** dipeptide via its thiol group, as illustrated in Figures 3.5 and 4.4. A His₆-Obal C1141A construct was generated to act as a negative control for the experiment using Gibson Assembly® (Gibson *et al.*, 2009 - Materials and methods section 2.4.6).

Unfortunately, a similar experiment was published in which TE-mediated **1** β -lactone ring formation was biochemically characterised (Schaffer *et al.*, 2017).

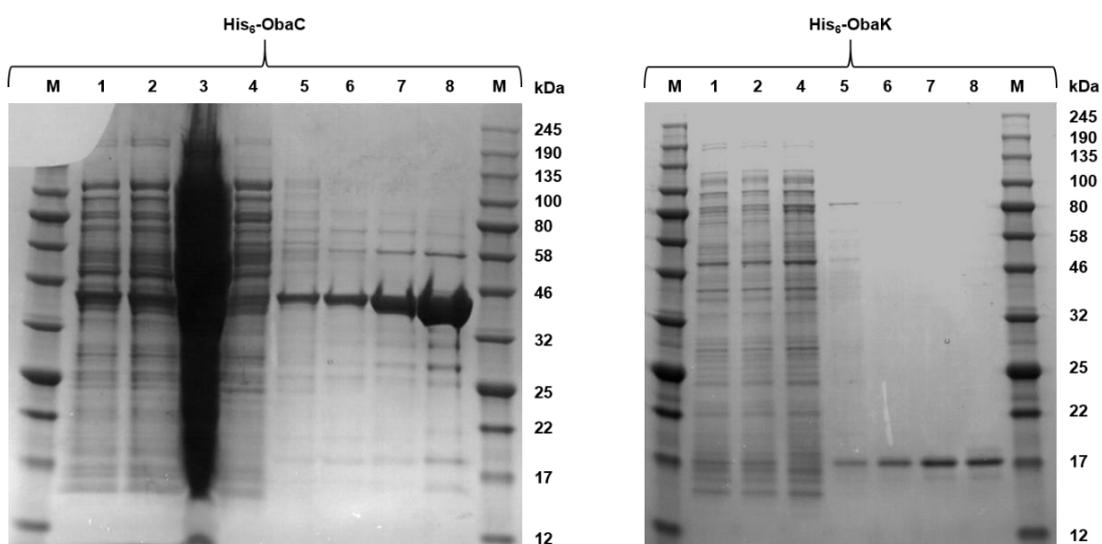


Figure 7.5. 12 % SDS PAGE gel of His₆-ObaC and His₆-ObaK purification steps. 1 = Cell lysate, 2 = Chitin wash eluent, 3 = 0 mM imidazole wash. 4 = 10 mM imidazole wash eluent, 5 = 20 mM imidazole wash eluent, 6 = 30 mM imidazole wash eluent, 7 = 50 mM imidazole wash eluent, 8 = 250 mM imidazole wash eluent, and M = Colour Prestained Protein Standard, Broad Range (NEB). The expected molecular weights for His₆-ObaC and His₆-ObaK are 38.2 and 12.2 kDa, respectively.

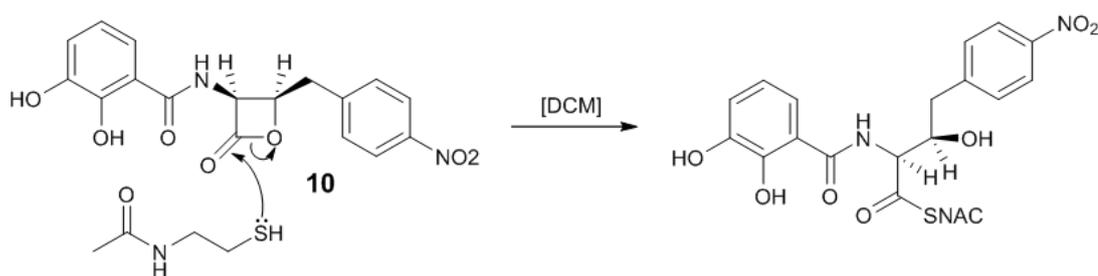


Figure 7.6. Illustration of obafluorin (1)-SNAC formation.

This was achieved by incubating WT *holo*-ObaK and WT (C1141), C1141S and C1141A variants of *holo*-Obal with **2**, **6** and ATP. Only the WT form of Obal could catalyse β -lactone ring formation. **10** was detected in experiments with the C1141 copy, but was also detected following incubation of **2** and **6** with the C1141S Obal variant. Neither **1** nor **10** were detected following the C1141A Obal substitution, consistent with the TE being responsible for β -lactone ring formation. Schaffer *et al.* also biochemically characterised ObaC by introducing it into a combined reaction with ObaG and ObaH using **7** and L-threonine as substrates. These three enzymes were able to catalyse the production of **2**, providing further evidence that **7** is the true substrate of ObaC.

7.4 Discussion

The NRPS Obal can be purified as a soluble N-terminally hexahistidine-tagged protein in its *apo*-form. *Apo*-His₆-Obal is active *in vitro* and could be used in a hydroxylamine-trapping assay for A domain substrate specificity. The assay revealed **2** to be the preferred substrate for the A₂ domain as expected. Intriguingly, L-serine and L-tryptophan were also accepted, though they generated significantly less signal than **2**, indicating some degree of substrate promiscuity in the Obal A₂ domain. When comparing the Obal A₂ amino acid residues that determine substrate specificity (Stachelhaus *et al.*, 1999) in Obal A₂ against those for L-phenylalanine and L-threonine-specific A domains (structurally similar substrate and substrate predicted by NRPSpredictor2 for Obal A₂, respectively), the Obal A₂ domain contains several important differences (Schaffer *et al.*, 2017 - Figure 7.7). Compared to phenylalanine-activating domains, Obal A₂ contains a glycine, not a threonine, at position 278, which would extend the substrate pocket to accommodate the C₄ side-chain of **2**. When compared to threonine-activating A domains, isoleucine at position 299 is a cysteine in Obal A₂, possibly providing a hydrogen-bonding partner for the aromatic nitro group of **2**, as well as facilitating interaction with the benzyl ring. Alanine is present instead of phenylalanine at position 236, creating space to accommodate the aromatic ring for **2** in place of the phenylalanine phenyl group. This unusual amino acid specificity conferring code could enhance future A domain substrate predictions based on the ten letter Stachelhaus code.

BA and 3-HBA were tolerated as substrates for the A₁ domain, complementing mutasynthetic results described in Chapter 8. *In vitro* A domain substrate promiscuity has been reported in several pseudomonad NRPS pathways (Gerard *et al.*, 1997; Wuest *et al.*, 2009). Such inherent plasticity is presumably an adaptive feature that

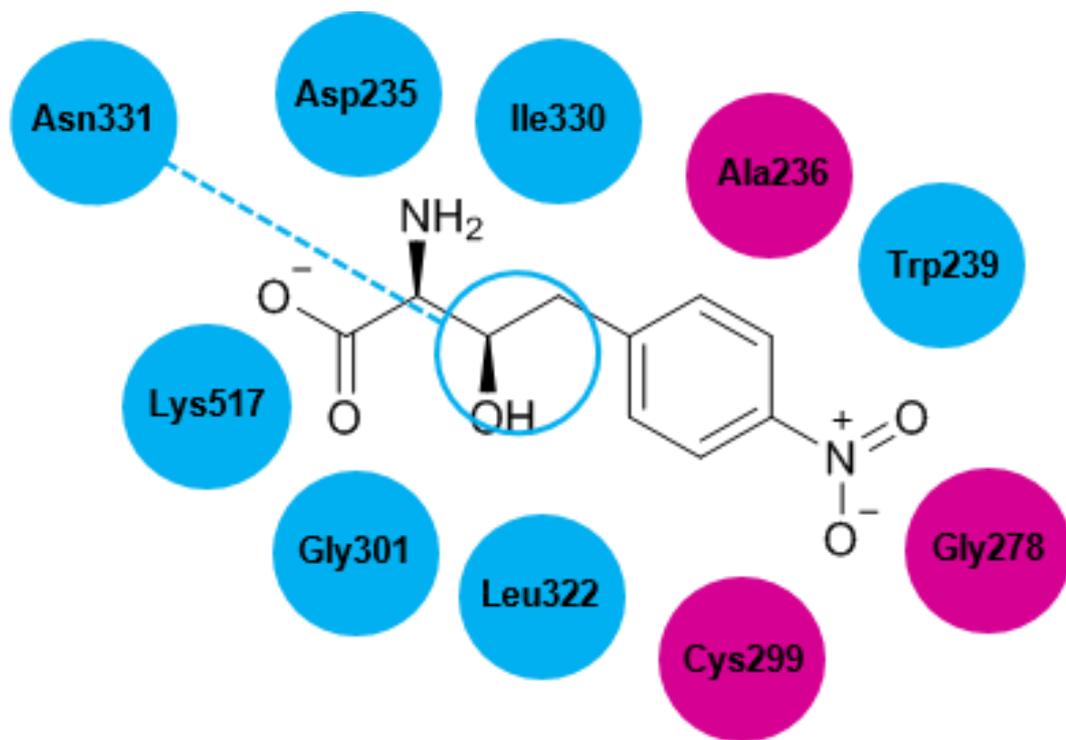


Figure 7.7. Illustration of the Obal A₂ domain substrate-specifying residues. Residues that differ from L-phenylalanine- or L-threonine-specific A domains are indicated in pink (Stachelhaus et al., 1999).

allows NP-producers to generate novel structures from single pathways and respond to new environmental stimuli. In support of this, single point mutations are sufficient to completely alter the substrate specificity of NRPSs, as has been shown in the biosynthesis of novel analogues of calcium-dependent antibiotic (Thirlway *et al.*, 2012), luminmide (Bian *et al.*, 2015) and gramicidin (Kries *et al.*, 2014). Despite apparent flexibility of both A domains in Obal, no congeners of **1** were detected in WT *P. fluorescens* ATCC 39502 production cultures, although such molecules might be susceptible to proteolytic degradation. More likely however, is that the C₂ domain exerts a 'gatekeeper' function with regards to the selection of substrates for carrier protein acylation and subsequent amide bond formation. This has been demonstrated for the cyanobacterial hepatotoxin microcystin, in which the downstream C domains are responsible for rendering modules monospecific for amino acid substrates rather than the A domains (Meyer *et al.*, 2016).

The MbtH-like domain could also be critical for determining substrate specificity of the Obal domains. MbtH-like proteins/domains share a conserved fold comprising an α -helix that packs onto a 3-stranded antiparallel β -sheet. Multiple sequence alignments have identified a sequence of 15 invariant, mostly hydrophobic, amino acids including three characteristic tryptophan residues that are common to all MbtH-like proteins (Baltz *et al.*, 2011). Structural studies on SlgN1, an A-MbtH didomain protein involved in streptolydigin biosynthesis (Herbst *et al.*, 2013), showed that these hydrophobic residues are mostly clustered on the face of its MbtH-like domain and pack against a small α - β docking motif on the A domain. This docking domain terminates in an alanine residue, the side chain of which projects into a hydrophobic cleft in the MbtH-like domain. Given that Obal can be purified in an active form and A domain activity can be assayed, future work will include generating point mutants in putative key MbtH-like domain residues to determine their impact on A domain activity both *in vivo* and *in vitro*. In parallel, we have begun a collaboration with the Prof. Changjiang Dong at the UEA to crystallise Obal. If this is possible, it should allow the characterisation of A domain/MbtH-like interactions in Obal at the structural level.

During the preparation of experiments to characterise TE-mediated cyclisation and to reconstitute the biosynthetic pathway for **1** *in vitro*, another group achieved these results (Schaffer *et al.*, 2017). Schaffer *et al.* showed that alteration of the *holo*-Obal TE active site cysteine to a neutral alanine residue (C1141A) abolished the ability of Obal and ObaK to assemble **1** *in vitro*. This is likely because the **1** dipeptide thioester is stalled on the PCP domain, as acyl transfer to the TE is no longer possible.

Intriguingly, a C1141S variant did not lead to **1** production, but low levels of **10** were detected, which indicates that this amino acid substitution converts Obal to a classic hydrolase, leading to the release of a linear dipeptide product. The difference between the C1141 and C1141S variants is likely due to the relative weakness of the C-S bond of the thioester compared to the C-O bond of the oxo-ester, which is necessary in order to make strained β -lactone ring formation thermodynamically favourable (McGrath and Raines, 2011). The higher ground-state energy of the thioester is also predicted to increase the rate of alcohol addition to the carbonyl group, which is likely to be the rate-limiting step. These thermodynamic requirements presumably are not met by the oxo-ester (Schaffer *et al.*, 2017). An active site cysteine residue was also recently reported in the SulM TE domain which could similarly be necessary for β -lactam ring formation and product release during the biosynthesis of sulfazecin (Li *et al.*, 2017).

Given that **1** assembly by Obalk, and the biosynthesis of **2** from **7** and L-threonine have both been reconstituted *in vitro*, further biochemical experiments regarding the biosynthesis of **1** were abandoned to instead focus on the biological significance of the β -lactone.

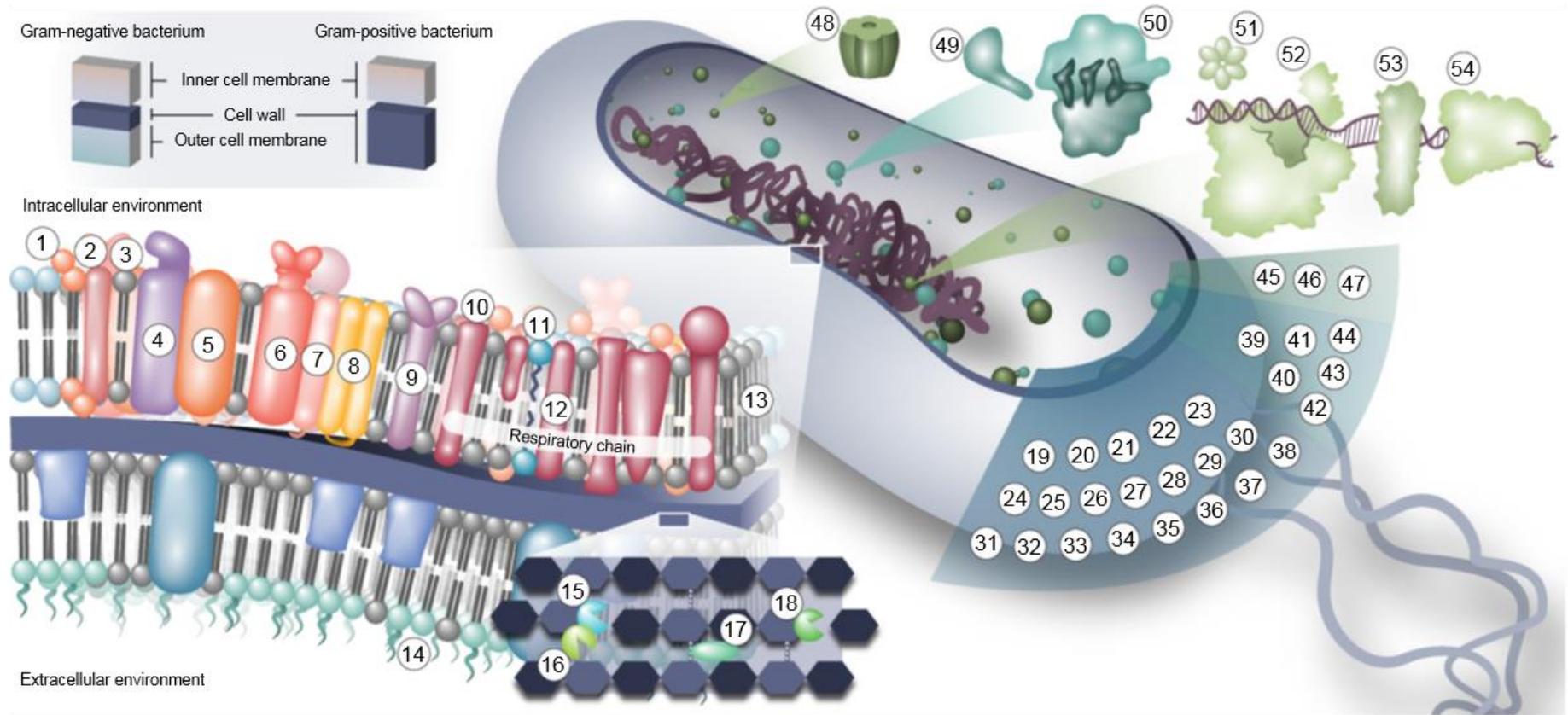
Chapter 8:
Characterising the putative
resistance gene and mode of
action for obafluorin (1)

Chapter 8: Characterising the putative resistance gene and mode of action for obafluorin (1)

8.1 Introduction

NPs encompass an astonishing variety of chemical structures, which allow them to mediate specific interactions with biological targets. Though many different selective advantages have been shown to be conferred by NPs (Traxler and Kolter, 2015), the majority are believed to act as chemical weapons that allow the producer to out-compete competitors in resource-limited environments. Bacterial and fungal NP targets are often sufficiently structurally different to their mammalian target equivalents that they will not interact with the latter, and so NPs have been employed extensively by humans as a source of antibiotics. Following hundreds of millions of years of evolution and diversification, it is perhaps unsurprising that for almost every known druggable target, there is at least one NP that targets it (Johnston *et al.*, 2016), a factor which has contributed significantly to their re-emergence in the genome sequencing era as a promising source of novel drug leads. In excess of 50 distinct molecular targets have been identified for antibacterial NPs (Figure 8.1), and the majority of these are inhibitors of three essential cellular processes: 1) Cell wall biosynthesis; 2) protein synthesis; and 3) DNA replication and repair (Walsh, 2000).

NP discovery today is driven by the rapid development of resistance to widely used antibiotics and the need to identify new ones to replace them. Antibiotics impose an extremely high selection pressure for the development of resistance, which is facilitated by the large number of bacteria involved in a given infection cycle, relatively high rates of intrinsic mutation, and the rapid spread of genes encoding resistance via MGEs and HGT (Walsh, 2000). Three principal mechanisms by which resistance is achieved are: 1) the use of drug efflux pump proteins to keep intracellular antibiotic concentration at ineffectual levels, as has been observed for tetracyclines and erythromycins (Ross *et al.*, 1990; Paulsen *et al.*, 1996); 2) the enzymatic modification of the antibiotic chemical structure to destroy its 'warhead'. This is exemplified by β -lactamases which cleave the pharmacophore β -lactam antibiotics before they can interact covalently with their PBP targets in the cytoplasmic membrane (Philippon *et al.*, 1985); and 3) the modification of antibiotic target structure to prevent or minimise antibiotic interaction, as illustrated by the *vanHAX* genes in enterococcal bacteria. Their gene products lead to the incorporation of D-alanine-D-lactate units into peptidoglycan, as opposed to the normal metabolite D-alanine-D-alanine, resulting in



a 1,000-fold drop in vancomycin binding affinity whilst having no detrimental impact on crosslinking efficiency (Bugg *et al.*, 1991; Walsh *et al.*, 1996). In order to mitigate the threat of antibiotic resistance, it is of critical importance to identify novel NPs with distinctly different modes of action to those of known antibiotics.

As mentioned in the introduction, **1** displayed only weak antibiotic activity against a range of different bacteria in early bioactivity experiments, but results were likely to be confounded by the inherent lability of the molecule and propensity for the β -lactone ring to be hydrolysed (Wells *et al.*, 1982; Tymiak *et al.*, 1985). However, in live-animal models, **1** was shown to protect mice challenged with *Streptococcus pyogenes* (ED_{50} = 50 mg/kg by systemic administration) and thus does display therapeutic potential. This, together with the observation that **1** also causes elongation of *E. coli* cells when administered at sub-lethal doses indicates that the molecule acts in a specific manner rather than as a general acylating agent. Given that β -lactone NPs are proven inhibitors of a diverse range of biological targets (De Pascale *et al.*, 2011), the mode of action for **1** warranted further investigation.

One fundamental condition of being able to synthesise a cytotoxic molecule is that a producing organism must have some mechanism of protecting itself from its own NP's adverse effects. This means that mechanisms of resistance against each antimicrobial NP inevitably already exist in nature, and resistance has been shown to be an ancient phenomenon (D'Costa *et al.*, 2011). As a result, loci encoding resistance determinants are commonly found adjacent to the biosynthetic genes required for the production of an antibiotic metabolite (Cundliffe and Demain, 2010), thereby facilitating their co-inheritance and utility in recipient cells (Osborn, 2010). This property has been successfully employed in genome-mining efforts to identify novel glycopeptide antibiotic BGCs in actinomycete strains resistant to vancomycin (Thaker *et al.*, 2013), and in this Chapter the application of similar logic to identify a putative resistance gene associated with the BGC for **1** is reported. Preliminary bioassay data collected by an MSc student under my supervision, Ms. Eposi Enjema Carine Solange (Carine), are also presented which validate the biological target of **1**. Experiments Carine performed to explore the metal-binding capacity of **1** are also described. Novel congeners of **1** were generated by mutasynthesis and will form the basis of future work to fully elucidate the mechanism of **1** bioactivity.

8.2 Identifying the putative resistance gene in *oba* BGC

When the project began, only one other 1-like BGC had been deposited in GenBank. This locus was found in the genome of *Chitiniphilus shinanonensis* DSM 23277 (NZ_KB895358.1), an environmental chitin-degrading β -proteobacterium (Figure 8.2). All *oba* genes have direct homologues in this cluster (except for *obaA* (*luxI* homologue) which is absent in *C. shinanonensis*) which all share between 60-99% identity at the amino acid level. *obaD* and *obaE* ADC synthase component homologues appear to be fused in one single gene in *C. shinanonensis*, analogously to *cmIB* and *papA* in the chloramphenicol and pristinamycin BGCs, respectively (Blanc *et al.*, 1997; He *et al.*, 2001). Genes flanking either cluster appear to be markedly different although a threonyl-tRNA synthetase (ThrRS) homologue is also present in both BGCs. The significance of this could not be determined given the low numbers of BGCs for comparison, uncertainty about how the *obaB* homologue in *C. shinanonensis* came to be located at the opposite end of the cluster compared to the ATCC 39502 BGC, and the lack of knowledge as to whether the BGC is even functional in this bacterium.

During the course of the project, several other putative *oba* BGCs were identified in genome sequences deposited in GenBank, allowing a more robust comparison to be made (Figure 8.2). A number of *Burkholderia* spp. (*B. diffusa* - NZ_LOTC01000033.1, *B. stagnalis* - NZ_LPGD01000044.1, *B. territorii* - NZ_LOSY01000044.1, and *B. ubonensis* - NZ_CP013463.1) were identified as carrying *oba*-like BGCs, albeit in a slightly rearranged form. *obaA* and *obaB* homologues in *Burkholderia* spp. are situated at the opposite end of the BGC to ATCC 39502, and the *obaM* homologue is located adjacent to an *obaC* homologue. Intriguingly, the A domain for **6** is encoded independently of the rest of the NRPS machinery in these clusters and *obaF* is split into independent chorismate mutase and prephenate dehydrogenase-encoding genes.

Two further *P. fluorescens* strains were identified in an analysis of *Pseudomonas* genomic diversity in take-all infected wheat fields (Mauchline *et al.*, 2015), which also comprise putative *oba* BGCs. These are 98-100% identical to that in ATCC 39502 at the amino acid level, and also comprise *orfs12345678*. They only share homology as far as the lysine tRNA gene downstream of *obaA-N* in ATCC 39502, however (Figure 8.2). This is consistent with the lysine tRNA gene being the likely boundary of the *oba* BGC, having been previously identified as a hot-spot for genomic island integration (Williams, 2002; Mavrodi *et al.*, 2009). Dr. Jacob Malone kindly provided both

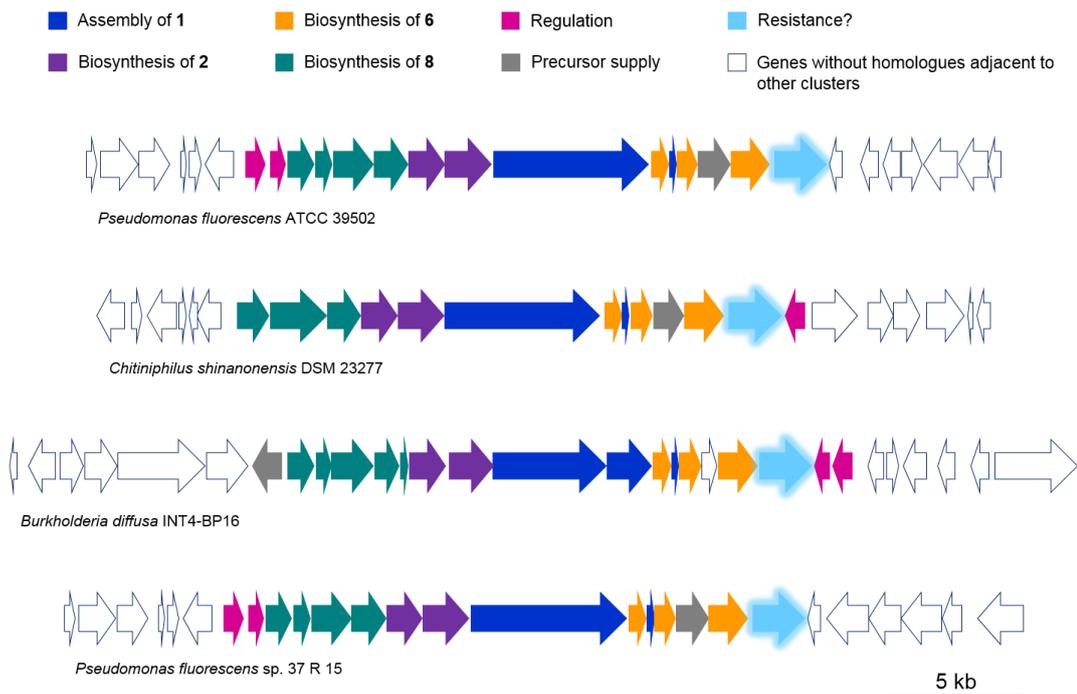


Figure 8.2. Illustration of putative *oba* BGCs identified in genomic sequence data uploaded to GenBank. A key is provided above that is consistent with Figures 3.5 and 4.4. Putative 1 resistance genes are highlighted in sky blue.

1-producing *P. fluorescens* strains (34 E 7 - NZ_CVTX01000156.1; and 37 R 15 - NZ_CVTV01000010.1), and when grown in OPM both produce **1** and cultures go purple.

Critically, in addition to characterised and predicted regulatory and biosynthetic genes, a PCS encoding a ThrRS was always found associated with putative *oba* BGCs, identifying it as the best candidate for a self-resistance gene. ThrRSs are class II aminoacyl-tRNA synthetases (aaRS) that are essential enzymes in protein synthesis that ligate L-threonine to its cognate tRNA molecule (Sankaranarayanan *et al.*, 1999 – Figure 8.3). aaRSs are well-documented targets for antimicrobial NPs (Cochrane *et al.*, 2016) and borrelidin is a known example of a NP ThrRS-inhibitor (Fang *et al.*, 2015). Further *in silico* analysis identified an additional gene encoding a ThrRS located elsewhere in the ATCC 39502 genome (in addition to the putative *oba* cluster-associated ThrRS copy) which likely represents the essential primary housekeeping copy of the enzyme. The *oba* ThrRS copy should only be essential under **1**-producing conditions, and is the likely result of a duplication of a primary enzyme which has evolved to be resistant to **1**. *C. shinanonensis* DSM 23277, *Burkholderia* spp. encoding putative *oba* BGCs, and *P. fluorescens* sp. 37 E 15 also all encode only one other 'primary' ThrRS in their genomes in addition to *obaO* homologues.

8.3 Experimental validation of *obaO* as a resistance determinant for obafluorin (1**)**

8.3.1 Disruption of *obaO*

obaO was disrupted using pTS1 and Δ *obaO* production culture samples were analysed by HPLC (Figure 8.4a). The *obaO* deletion led to the abolition of **1** production and the loss of the characteristic purple phenotype (Figure 8.4b), consistent with a critical role in some element of **1** production. Unfortunately, initial attempts to complement this mutant genetically by introducing pJH10TS-*obaO* failed, however sequence data were consistent with a clean in-frame deletion that should not impede *obaJKLMN* transcription. Furthermore, characteristic shunt metabolites **4** and **5** also did not accumulate in the Δ *obaO* background, where they did in **6** biosynthesis mutants Δ *obaJ* (Figure 4.5) and Δ *obaL* (Figure 4.6), suggesting that the phenotype cannot be explained by polar effects of the deletion on **6** biosynthesis alone. However, further work is required to absolutely preclude possible polar effects of *obaO* deletion.

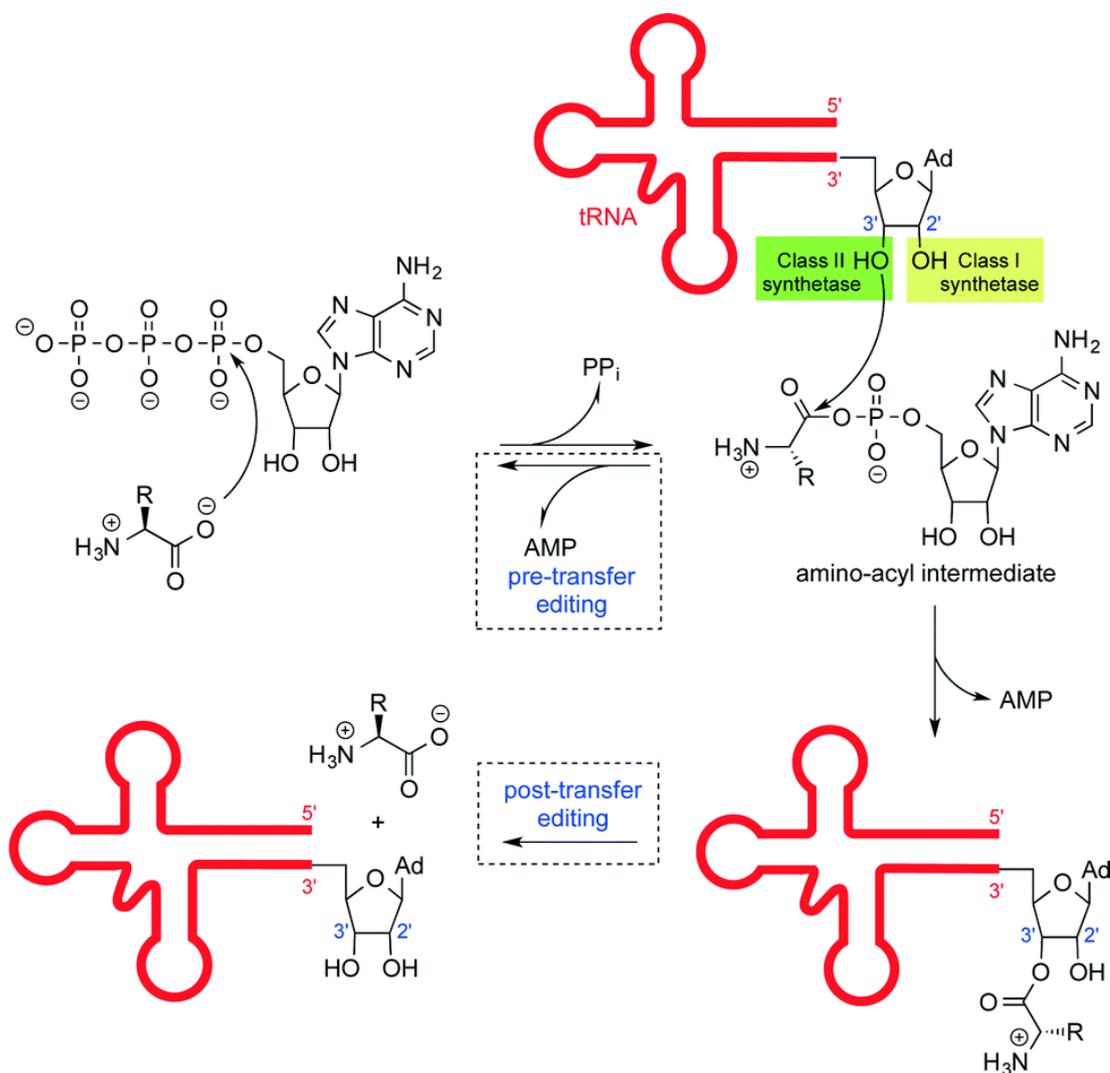


Figure 8.3. Mechanism by which aminoacyl tRNA synthetases (aaRSs) ligate amino acids to their cognate tRNAs. The amino acid first attacks ATP to yield an aminoacyl-adenylate with concomitant release of pyrophosphate. The amino acid is then transferred to the 2' (class I) or 3' (class II) hydroxyl of the 3' terminal adenosine residue of the cognate tRNA molecule to generate an aminoacyl-tRNA complex. Some aaRS enzymes are also capable of pre- or post-transfer editing. Cochrane et al. (2016) - Reproduced with permission from The Royal Society of Chemistry.

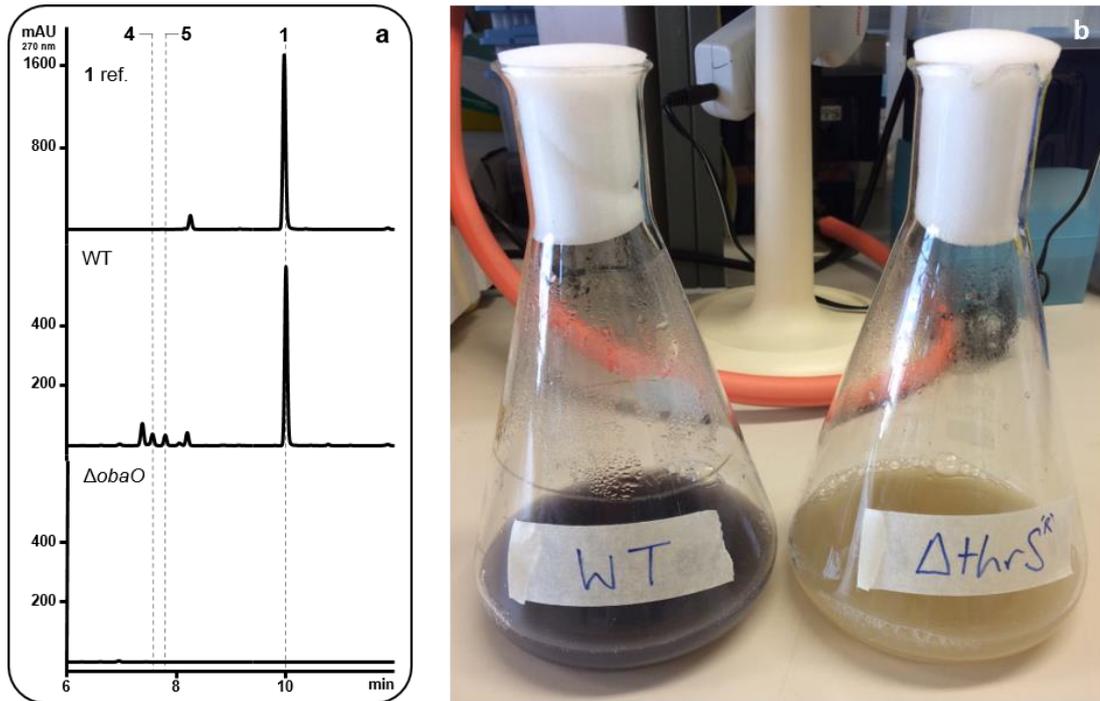


Figure 8.4. Mutagenesis HPLC data for $\Delta obaO$ and image of the corresponding production cultures. (a) HPLC profiles for *obaO* mutagenesis and complementation experiments at 270 nm. (b) WT (left) and $\Delta obaO$ (right) production cultures (*thrS*^R = *obaO*).

8.3.2 Assaying obafluorin (1) antibacterial activity

In order to determine whether *obaO* confers resistance to **1**, thereby confirming **1** as a ThrRS inhibitor, antibacterial activity assays were conducted to determine the minimum inhibitory concentration (MIC) of **1** against three different bioassay strains: 1) *E. coli* ATCC 25922, a strain traditionally used in antibiotic susceptibility experiments (Boyle *et al.*, 1973); 2) *E. coli* NR698 (Ruiz *et al.*, 2005), which carries an in-frame deletion in the *imp4213* gene which is responsible of outer-membrane lipopolysaccharide assembly (Ruiz *et al.*, 2006), thereby increasing antibiotic permeability into the cell; and 3) *B. subtilis* EC 1524 (O'Rourke *et al.*, 2017), as a representative of Gram-positive bacteria. Given that **1** is a substrate for β -lactamases (Wells *et al.*, 1984), *Pseudomonas* spp. were not assayed as the genomes of many encode one or more of these enzymes (Livermoore, 2002; Michaux *et al.*, 2008); five putative β -lactamases are encoded in the ATCC 39502 genome. All antibiotic susceptibility assays described were performed by Carine and are described in Materials and methods section 2.9.

8.3.2.1 Broth microdilution assays

Initially, test strains were challenged with **1** over a range of concentrations between 0.015-2000 $\mu\text{g/mL}$ in a liquid culture, microtitre plate-based format. MICs were determined to be 2000 $\mu\text{g/mL}$ for both *E. coli* ATCC 25922 and *B. subtilis* EC 1524, and 1000 $\mu\text{g/mL}$ for *E. coli* NR698. No growth was observed in positive controls containing either 100 $\mu\text{g/mL}$ carbenicillin or 50 $\mu\text{g/mL}$ kanamycin. Whilst one would anticipate *E. coli* NR698 to be slightly more susceptible to **1** given the *imp4213* deletion, MIC values recorded were far in excess of what would be expected for a genuine antibiotic compound. However, these results were not consistent with previous MICs recorded from disc-diffusion based assays for **1** against several other *E. coli* strains (Table 8.1 - Wells *et al.*, 1984; Tymiak *et al.*, 1985). At concentrations where cell growth was unaffected, small films were observed at the bottom of the wells, suggesting possible biofilm formation. Many bacteria can encase themselves in matrices composed of polysaccharides and proteins known as biofilms, which limit antibiotic penetration and so confer resistance (Stewart and Costerton, 2001). To prevent cells from aggregating and forming biofilms, the experiment was repeated in a rotary shaker with 200 rpm agitation, but similar MIC values were recorded despite the disappearance of film-like structures.

Organism	SC strain no.	Zone of inhibition (mm) (Obafluorin (1) – 10 µg)
<i>Staphylococcus aureus</i>	1276	8.7
	2399	7.4
	2400	7.9
<i>Escherichia coli</i>	8294	-
	10857	10.5
	10896	12.8
	10909	13.4
<i>Enterobacter cloacae</i>	8236	9.3
<i>P. rettgeri</i>	8479	7.7
<i>P. aeruginosa</i>	9545	9.1

Table 8.1. Obafluorin (1) disc diffusion assay results from Wells et al. (1982). In these experiments inoculum was adjusted with McFarland 0.5 turbidity standard and tested on K10 agar (pH 6.7). 2 mg/mL Obafluorin (1) was dissolved in MeCN and 5 µL were applied to 6.3mm Whatmann #4 filter discs.

8.3.2.2 Disc diffusion assays

Given the unexpectedly high MIC values obtained in liquid culture, agar plate-based methods were subsequently employed to confirm the validity of these results. In initial experiments, **1** dissolved in 100% MeCN over the same range of concentrations as before, 5 µL was applied to paper discs before being placed on agar plates inoculated with one of the three WT test strains described. No measurable zone of bacterial growth inhibition was observed following incubation at any concentration of **1** tested, nor were they for the MeCN vehicle negative control. In contrast, halos of inhibition were clearly visible for 100 µg/mL carbenicillin and 50 µg/mL kanamycin positive controls. These data were consistent with the previous finding that **1** binds to paper discs, limiting its diffusion into agar (Tymiak *et al.*, 1985).

8.3.2.3 Agar diffusion assays

Finally, assays were performed in which 5 µg/mL of **1** was pipetted directly onto the agar surface. Zones of clearing could be clearly observed following direct application, allowing new MIC values to be determined: *B. subtilis* EC 1524 = 0.125 µg/mL; *E. coli* ATCC 39502 = 32 µg/mL; and *E. coli* NR698 = 4 µg/mL (Table 8.2). A negative control in which MeCN alone was applied resulted in no observable impact on bacterial

growth. These results indicate that contrary to previous findings (Wells *et al.*, 1984; Tymiak *et al.*, 1985), **1** has potent antibacterial activity against both Gram-positive and -negative bacteria.

Bioassay strain	Obafluorin MIC ($\mu\text{g/mL}$)
<i>B. subtilis</i> 1524	0.125
<i>E. coli</i> ATCC 25922	32
<i>E. coli</i> ATCC 25922 pJH10TS	32
<i>E. coli</i> ATCC 25922 pJH10TS- <i>obaO</i>	>2000
<i>E. coli</i> NR698	4
<i>E. coli</i> NR698 pJH10TS	0.5
<i>E. coli</i> NR698 pJH10TS- <i>obaO</i>	>2000

Table 8.2. Preliminary obafluorin (1) agar diffusion assay results.

pJH10TS and pJH10TS-*obaO* were subsequently transformed into both *E. coli* strains and the agar diffusion assay was repeated to determine whether any degree of resistance is conferred by the putative **1** resistance gene. Remarkably, in both strains the MICs were >2000 $\mu\text{g/mL}$, representing >60-fold and >4000-fold increases for ATCC 25922:pJH10TS-*obaO* and NR698:pJH10TS-*obaO* respectively, when compared to the same strains carrying the empty vector (Table 8.2). These results demonstrate that ObaO confers resistance to **1**, identifying **1** as a genuine ThrRS inhibitor. Whilst these are preliminary data, they are consistent with observations made during the development of the assay. Bioassays with both *E. coli* strains at higher concentrations will also be performed to determine the exact MIC values as these were outside the range of **1** concentrations used in these preliminary experiments.

8.4 Identifying the obafluorin (1) mechanism of ThrRS inhibition

8.4.1 *In silico* comparison of putative obafluorin (1)- ‘sensitive’ and ‘resistant’ ThrRS amino acid sequences

Given that the residues critical for ThrRS catalytic activity are well defined (Sankaranarayanan *et al.*, 1999; 2000 – Figure 8.5), an alignment was performed in order to identify potential nucleophilic amino acid β -lactone targets. Amino acid sequences of putative **1** resistance genes were aligned against putative ‘sensitive’ ThrRSs from the same strains, in addition to those for which structures are available on the PDB database (Appendix 1 – Supplementary Figure 6). This included the *Homo sapiens* ThrRS which is known to bind another natural ThrRS inhibitor, borrelidin (Fang *et al.*, 2015). Almost all key active site residues are highly conserved

across the different sequences except for an active site cysteine residue (position 480 in *E. coli* – 1QF6), which in putative **1** resistance ThrRSs is consistently substituted by a valine residue. Cysteine is extremely nucleophilic by virtue of its thiol group, and would represent an ideal candidate residue for attack of the **1** β -lactone ring. If this residue were the target for **1**, it would explain why a valine is present in putative resistance ThrRSs as it is a particularly non-polar amino acid. A lysine residue is present at this position in *S. aureus* and whilst not as nucleophilic as cysteine, could still conceivably attack the **1** β -lactone ring. Bioassays would determine whether **1** is active against *S. aureus*.

8.4.2 *In silico* comparison of *obaO* with other putative ThrRS-inhibiting NP resistance determinants

A putative ThrRS, OzmT, is also encoded in the oxazolomycin BGC, indicating that in addition to **1** and borrelidin, oxazolomycin might also be a ThrRS inhibitor (Zhao *et al.*, 2010). An alignment of putative resistant ThrRSs for **1**, borrelidin (BorO) and oxazolomycin (OzmT) revealed that the active site residues in each are highly conserved (Appendix 1 -Supplementary Figure 7). Like ObaO, both BorO and OzmT possess alternative amino acids to the C480 residue in *E. coli* ThrRS. Borrelidin is known to occupy the active site of human ThrRS by forming hydrogen bonds with a number of active site residues to exclude enzyme substrates. In BorO an alanine is in place of this cysteine residue, but this does not participate in borrelidin binding, indicating that variation at this position is not necessarily associated with resistance. In OzmT, the residue at this position is a lysine, which whilst not as nucleophilic as cysteine, could still plausibly attack the β -lactone ring, though β -lactones have so far only been reported to react with cysteine, serine or threonine enzyme active site residues (Böttcher and Sieber, 2012). This could indicate an alternate mode of action for oxazolomycin.

The genome of *B. subtilis*, which was found to be highly sensitive to **1**, encodes two ThrRS genes, *thrS* and *thrZ* (Putzer *et al.*, 1990), which were included in the same alignment (Appendix 1 - Supplementary Figure 7). During vegetative growth, only the former is expressed, with *thrZ* expressed in response to decreases in intracellular ThrS concentration in a dose-compensatory manner (Putzer *et al.*, 1992). ThrS and ThrZ only share 57% sequence identity and possess lysine and cysteine residues respectively at the position equivalent to C480 in *E. coli* ThrRS. If both are expressed, **1** would have to be able to inhibit both copies in order to have its antibiotic effects, raising the question how it would mediate specific interactions with both proteins.

Alignments using putative resistance genes for BorO, OzmT and redundant *B. subtilis* ThrRS copies did not provide further support for the hypothesis that the ThrRS active site cysteine residue is the target of **1**. Biochemical experiments were performed to investigate interaction between **1** and ThrRS.

8.4.3 *E. coli* ThrRS obafluorin (1)-binding assay

In order to validate the hypothesis that **1** binds to C480 in native *E. coli* ThrRS, it was cloned for expression in pET28a(+) as an N-terminally hexahistidine-tagged protein (Figure 8.6). Once purified, His₆-ThrRS was exchanged into MES buffer (pH 6.0) and incubated with **1** before subsequent analysis by UPLC-HRMS (Materials and methods sections 2.7.7 and 2.10.2). All metabolomic analyses performed for these experiments were conducted by Dr. Daniel Heine and Dr. Gerhard Saalbach. As a negative control, Gibson Assembly® (Gibson *et al.*, 2009) was applied as for Obal TE mutations to create a construct with a C480V (TGC→GTG) substitution, mimicking the alteration observed in putative **1** resistance gene products (Materials and methods section 2.4.6). Both WT and *E. coli* ThrRS C480V equivalents appeared to bind **1**, as determined by a mass shifts of 358.2 and 358.6 Da for both proteins (**1** = 358.1 Da), respectively, following incubation with **1** (Figure 8.7). Samples were subsequently subjected to trypsin digest to identify specific residues modified by **1**. Analysis of these samples using Orbitrap Fusion (Thermo Fisher Scientific) and Synapt G2-Si (Waters) high-resolution mass spectrometers could not confirm the specific binding of **1** to the active site C480 residue, or to any other *E. coli* ThrRS active site residues. Further work is required to investigate **1** ThrRS-binding and is discussed in further detail in the discussion.

8.5 Investigating the metal-binding properties of obafluorin (1)

In addition to the β -lactone ring, **1** comprises two other structural elements, a catechol moiety and aromatic nitro group, both of which occur in a number of bioactive NPs (Crosa and Walsh, 2002; Winkler and Hertweck, 2007; Parry *et al.*, 2011). Whilst the role of nitro groups in bioactivity is poorly understood, the role of the catechol moiety in the coordination of ferric iron has been well characterised in numerous siderophore NPs (Hider and Kong, 2009). ThrRS requires a Zn²⁺ cofactor to coordinate L-threonine in its active site through the formation of a pentacoordinate intermediate with both the amino and side-chain hydroxyl groups (Sankaranarayanan *et al.*, 2000 – Figure 8.5).

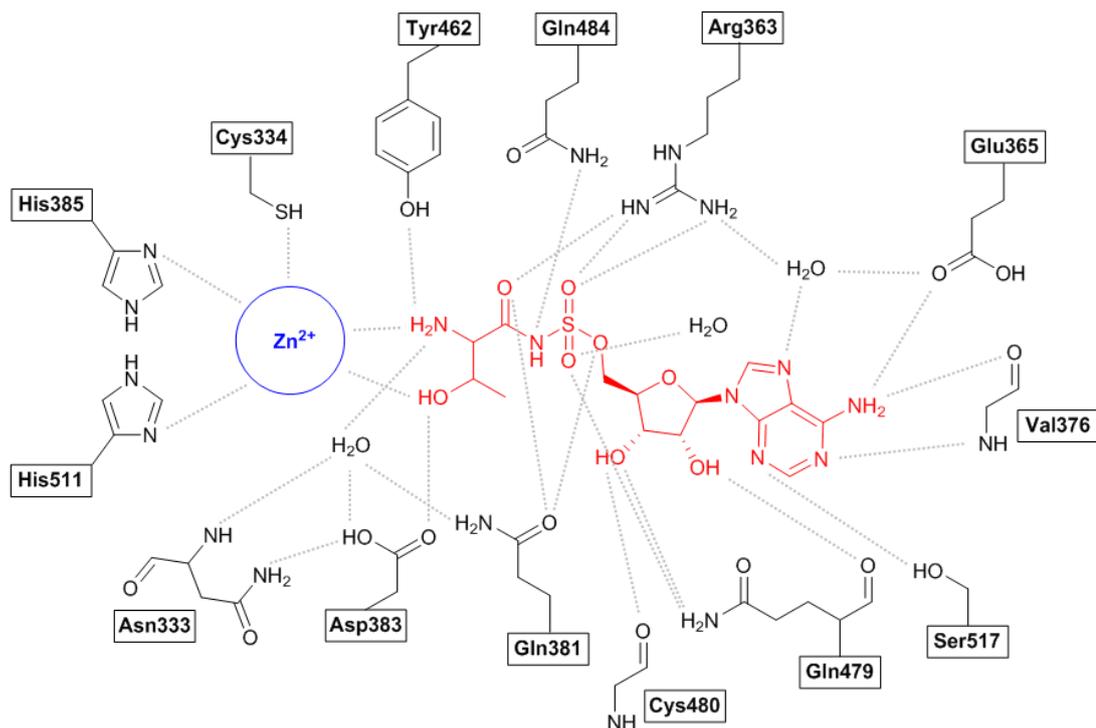


Figure 8.5. The interaction network of the Thr-AMS molecule with the Zn^{2+} cofactor and the ThrRS active site residues. Adapted from Sankaranarayanan et al. (2000) with permission from the Nature Publishing Group.

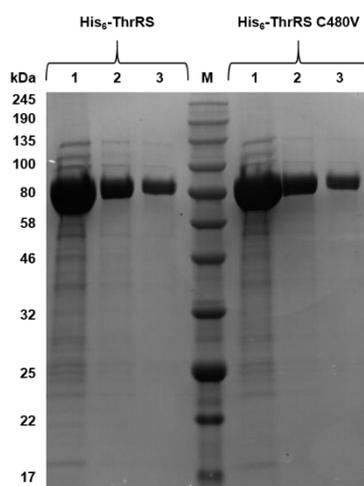


Figure 8.6. 12 % SDS PAGE gel of *E. coli* His₆-ThrRS and His₆-ThrRS C480V purified protein sample serial dilutions. 1 = 1/20, 2 = 1/100, 3 = 1/200, and M = Colour Prestained Protein Standard, Broad Range (NEB). The expected molecular weights for His₆-ThrRS and His₆-ThrRS (C480V) are both 76.2 kDa.

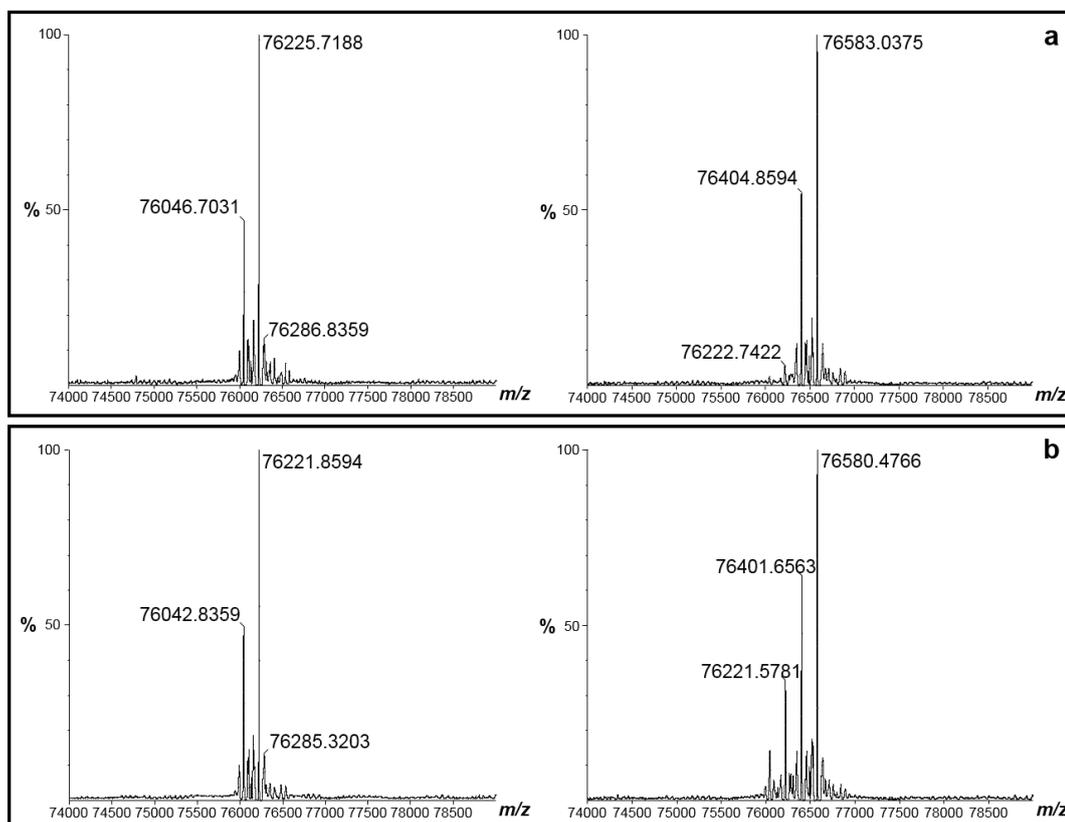


Figure 8.7. Mass spectra (74,000 – 79,000 m/z range) for *E. coli* ThrRS obafuorin (1)-binding assays. (a) *E. coli* His₆-ThrRS and (b) *E. coli* His₆-ThrRS C480V. Left = MeCN negative control, right = 1 test assay. The expected masses for *E. coli* WT and C480V His₆-ThrRS variant proteins are 76,177.62 and 76,173.61 Da, respectively. Discrepancies between expected masses and observed masses could possibly be attributed to formylation of the N-terminal methionine (28.01 m/z) and/or the presence of the zinc cofactor (65.38 m/z).

This raised the possibility that the **1** catechol might also participate in target interaction. The ability of **1** to bind Fe³⁺ and Zn²⁺ was therefore evaluated using two different assays and these experiments were also performed by Carine.

8.5.1 Chrome azurol S (CAS) assay to determine obafluorin (1) iron-binding potential

To explore the ability of **1** to bind ferric iron, a CAS assay was performed which assesses the ability of molecules to chelate Fe³⁺ bound to CAS, resulting in a colour change from blue through pink to orange (Schwyn and Neilands, 1987). Introduction of **1** to CAS-Fe³⁺ complex solution resulted in a colour change from blue to pink (Figure 8.8), indicating that **1** disrupts the complex to chelate iron and release the CAS dye. This was not observed in a negative control in which 100% MeCN was added instead, providing qualitative evidence that **1** is able to chelate ferric iron.

8.5.2 4-(2-pyridylazo)-resorcinol (PAR) assay to determine obafluorin (1) zinc-binding potential

A PAR assay (Bandara *et al.*, 2009) was performed to determine whether **1** could out-compete PAR to bind Zn²⁺ in solution. Addition of varying concentrations of **1** between 1 and 50 µM to PAR solution resulted in a decrease of the characteristic Zn(PAR)₂ complex absorbance peak at 493 nm (Figure 8.9), with greater decreases observed at higher concentrations of **1**. This suggests that **1** does indeed out-compete PAR for Zn²⁺. However, a concomitant increase in absorbance at 410 nm corresponding to the accumulation of unbound PAR was not observed. This was shown to occur in a similar experiment with holomycin (Chan *et al.*, 2017), and might suggest that PAR can react with another assay component, with **1** the most likely candidate as it is the variable component between our experiments and those previously published.

8.6 Generation of congeners of obafluorin (1) by mutasynthesis

To explore whether the catechol moiety is significant in the interaction of **1** with ThrRS, we desired access to congeners of **1** in which the catechol moiety was altered such that metal-binding affinity would be reduced or abolished. These congeners could be applied in bioassay and ThrRS-binding experiments to see whether such alterations impact the ability of the molecule to inhibit bacterial growth or to bind its putative ThrRS target, respectively.



Figure 8.8. Preliminary qualitative liquid CAS assay results. Left = CAS + MeCN negative control, Right = CAS + 1. The pink colour indicates that 1 is outcompeting CAS for Fe^{3+} , leading to dye release and colour change.

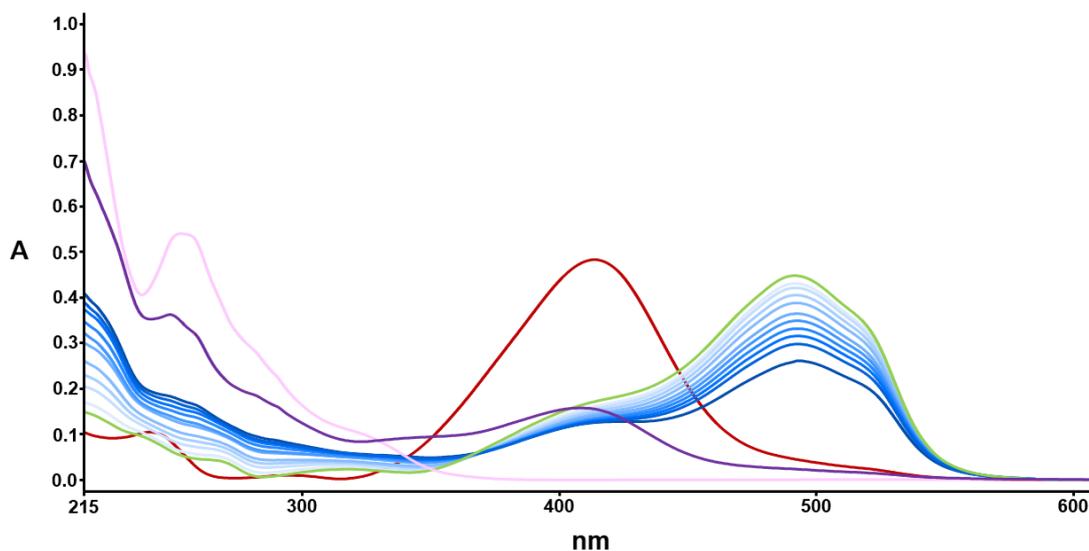


Figure 8.9. Preliminary PAR assay spectrophotometric data. Red = PAR only, Pink = 1 dissolved in MeCN only, Green = MeCN vehicle negative control, Blue = Test assays with incremental ($1 \mu\text{M}$) increases between 1(light)-10(dark) μM 1 to $\text{Zn}(\text{PAR})_2$ complex solution, and Purple = Test assay with $50 \mu\text{M}$ 1 added to the $\text{Zn}(\text{PAR})_2$ complex solution. Higher concentrations of 1 cause greater reductions of $\text{Zn}(\text{PAR})_2$ complex, indicating that it disrupts it by chelating Zn^{2+} .

To achieve this, *ΔobaL* strain production cultures were fed alternative benzyl carboxylic acids to see whether these would be accepted as alternative substrates to **6**, as illustrated in Figure 8.10. If successful, this would generate congeners with reduced metal-binding capacities which could be applied in both biochemical ThrRS-binding experiments and bioassays to compare their antibiotic activities relative to **1**. Benzoic acid (BA), salicylic acid (SAL), 3-hydroxybenzoic acid (3-HBA), 4-hydroxybenzoic acid (4-HBA), 3,5-dihydroxybenzoic acid, 2,3,4-trihydroxybenzoic acid, gallic acid and anthranilic acid (ANA) were all fed exogenously to *ΔobaL* production cultures to a final concentration of 0.4 mM (ANA was fed to a final concentration of 1 mM). Whilst none of the *ΔobaL* cultures fed alternate benzyl carboxylic acids displayed the purple colour associated with **1** (Figure 8.11), HPLC analysis revealed the generation of novel peaks in cultures fed BA, SAL, 3-HBA and ANA (Figure 8.12), suggesting the successful production of novel **1** congeners. In all cases, **1** production was no longer observed, but shunt metabolites **4** and **5** are still seen to accumulate, indicating that **2** biosynthesis is unaffected. Titres of **4** and **5** were very low in ANA and SAL-fed cultures, reflecting the fact that cultures grew poorly and titres simply did not reach observable levels. ANA was fed at lower concentrations (0.2-0.5 mM) but no change in peak areas were observed under these conditions, despite noticeably improved culture growth.

Culture samples were further analysed by HR-LCMS and accurate masses were detected for each of the anticipated **1** congeners (Figures 8.13-8.15), in addition to their hydrolysed, ring-open equivalents, in all cases except for ANA, in which the corresponding congener masses could not be detected. This could be because the ANA congener ionises poorly, and/or because of poor growth of cultures and low metabolite titres. These results match with hydroxylamine-trapping assay data probing A-domain substrate specificity which showed that both BA and 3-HBA are accepted as substrates by the Obal A₍₁₎ domain (Figure 7.3). In those experiments, BA appears to be better accepted relative to 3-HBA, and this reflected in the titres of their respective **1** congeners, based on integrated peak areas at 270 nm (Figures 8.13 and 8.15). Mutasynthesis experiments will be scaled up to generate sufficient material for NMR analysis to verify the structural identity of these congeners of **1**, particularly as SAL and 3-HBA congeners have the same expected masses. Material generated can then be used in biochemical and bioassay experiments to elucidate the **1** mode of action.

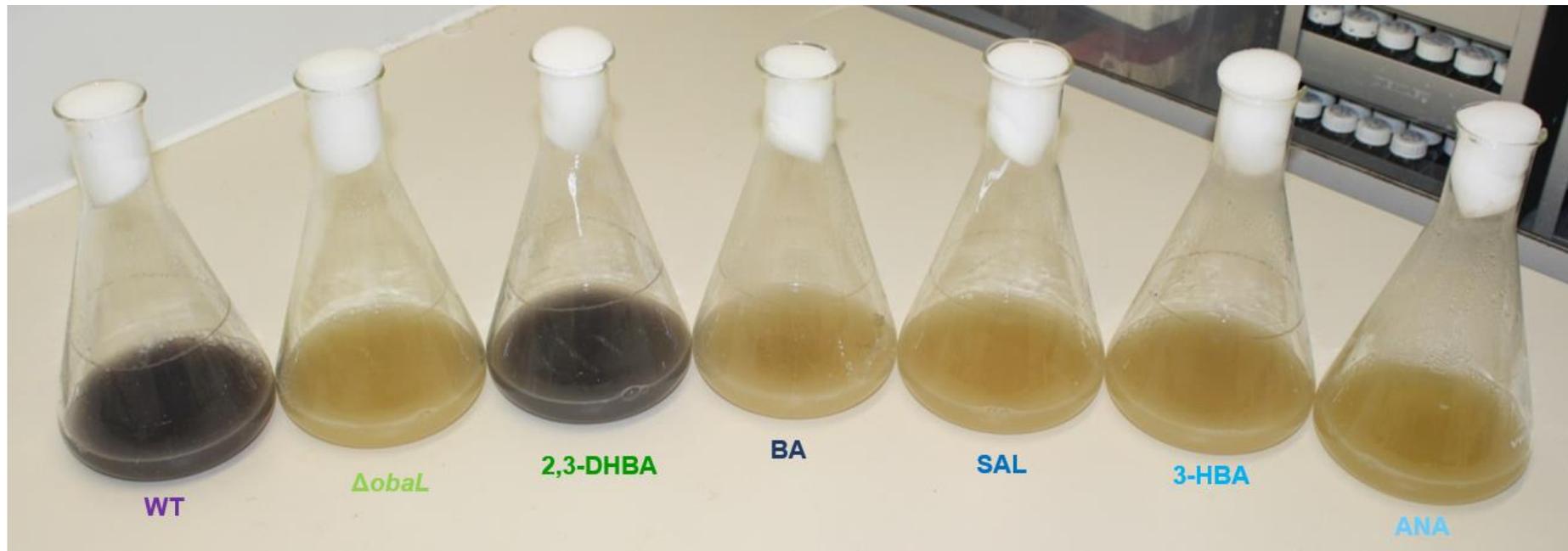


Figure 8.11. Image of mutasynthesis experiment production cultures. Left-Right: WT fed DMSO, Δ obaL fed DMSO, Δ obaL fed 6 (0.2 mM), Δ obaL fed BA (0.4 mM), Δ obaL fed SAL (0.4 mM), Δ obaL fed 3-HBA (0.4 mM) and Δ obaL fed ANA (1 mM).

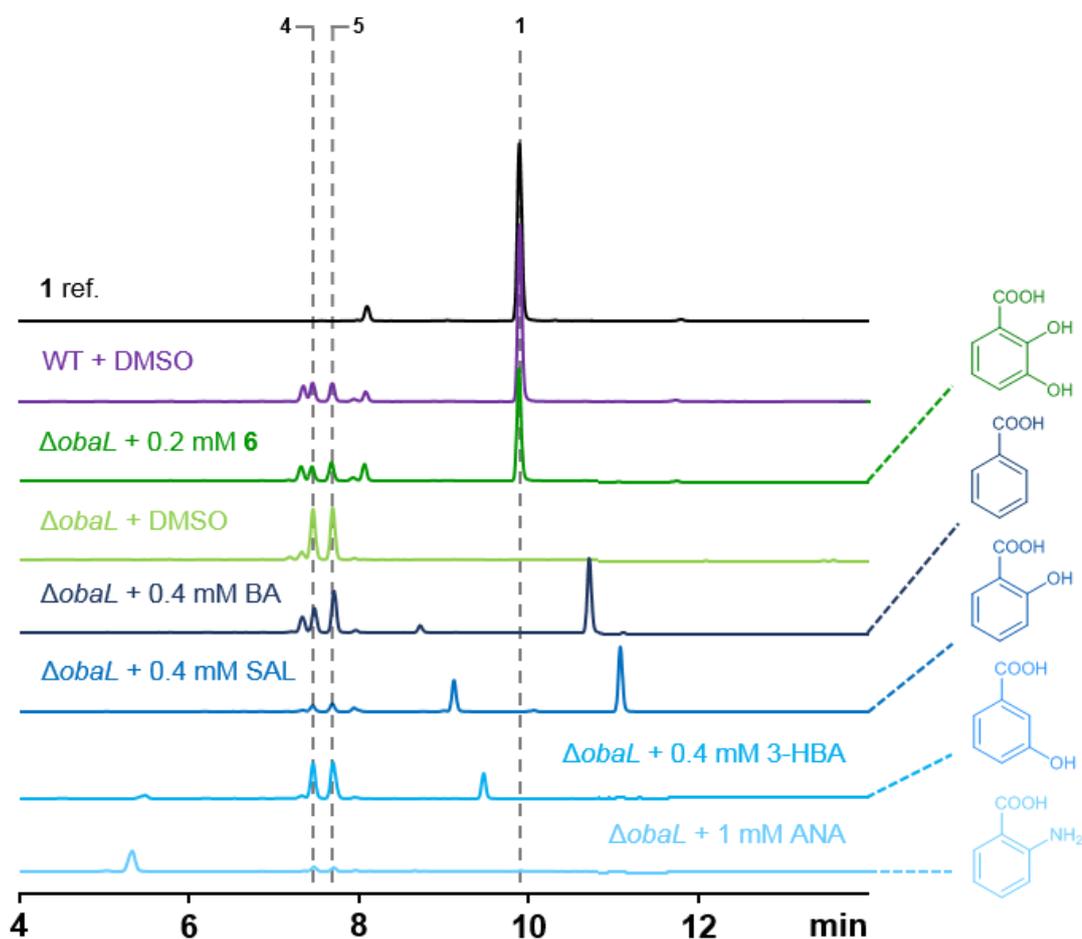


Figure 8.12. HPLC data for extracts of mutasynthesis experiment production cultures. Samples correspond to the cultures in Figure 8.11. Novel peaks can be seen in each of the test conditions indicating the biosynthesis of novel congeners of 1. Benzyl carboxylic acids fed are indicated for each trace and colours used coordinate with Figure 8.11. Samples here were analysed on the Agilent 1100 series HPLC. Subsequent analysis (Figures 8.13-15) was performed on a Nexera X2 (LC-30AD) system connected to a LCMS-IT-TOF Liquid Chromatograph mass spectrometer (Shimadzu) using an identical programme and column to collect MS data for each sample.

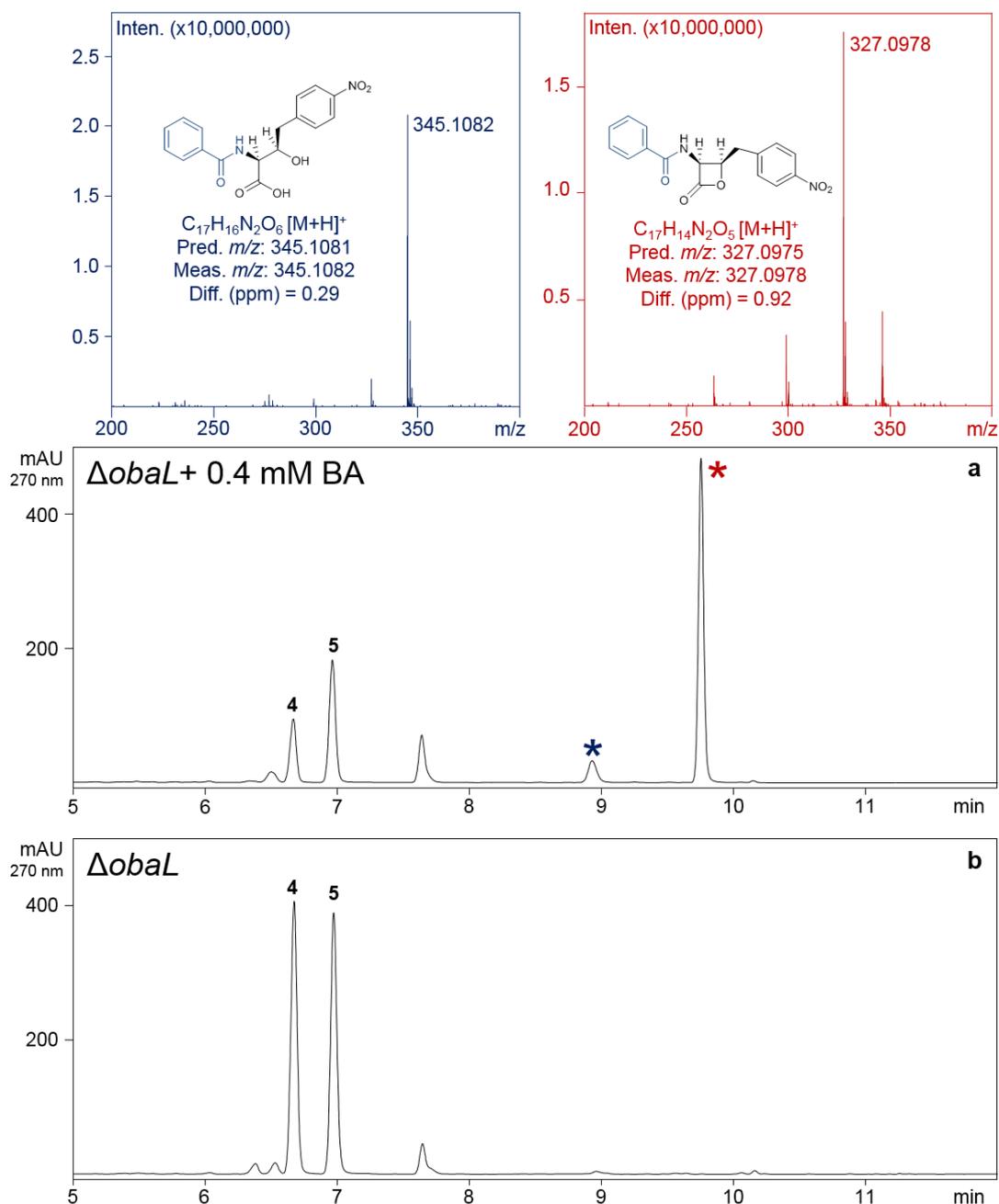


Figure 8.13. HPLC and MS data for Δ obaL fed 0.4 mM benzoic acid (BA). (a) UV chromatogram for a Δ obaL production culture fed BA to a final concentration of 0.4 mM. Extracted mass spectra for peaks identified in the BA-fed culture UV chromatogram with * are included above: Red = BA congener of 1 (Calculated for $C_{17}H_{14}N_2O_5 [M+H]^+$ - Pred. m/z: 327.0975, Meas. m/z: 327.0978, Diff. (ppm) = 0.92); Blue = Hydrolysed form of the BA congener of 1 (Calculated for $C_{17}H_{16}N_2O_6 [M+H]^+$ - 345.1081, Meas. m/z: 345.1082, Diff. (ppm) = 0.29). (b) UV chromatogram of a Δ obaL production culture fed DMSO as a negative control.

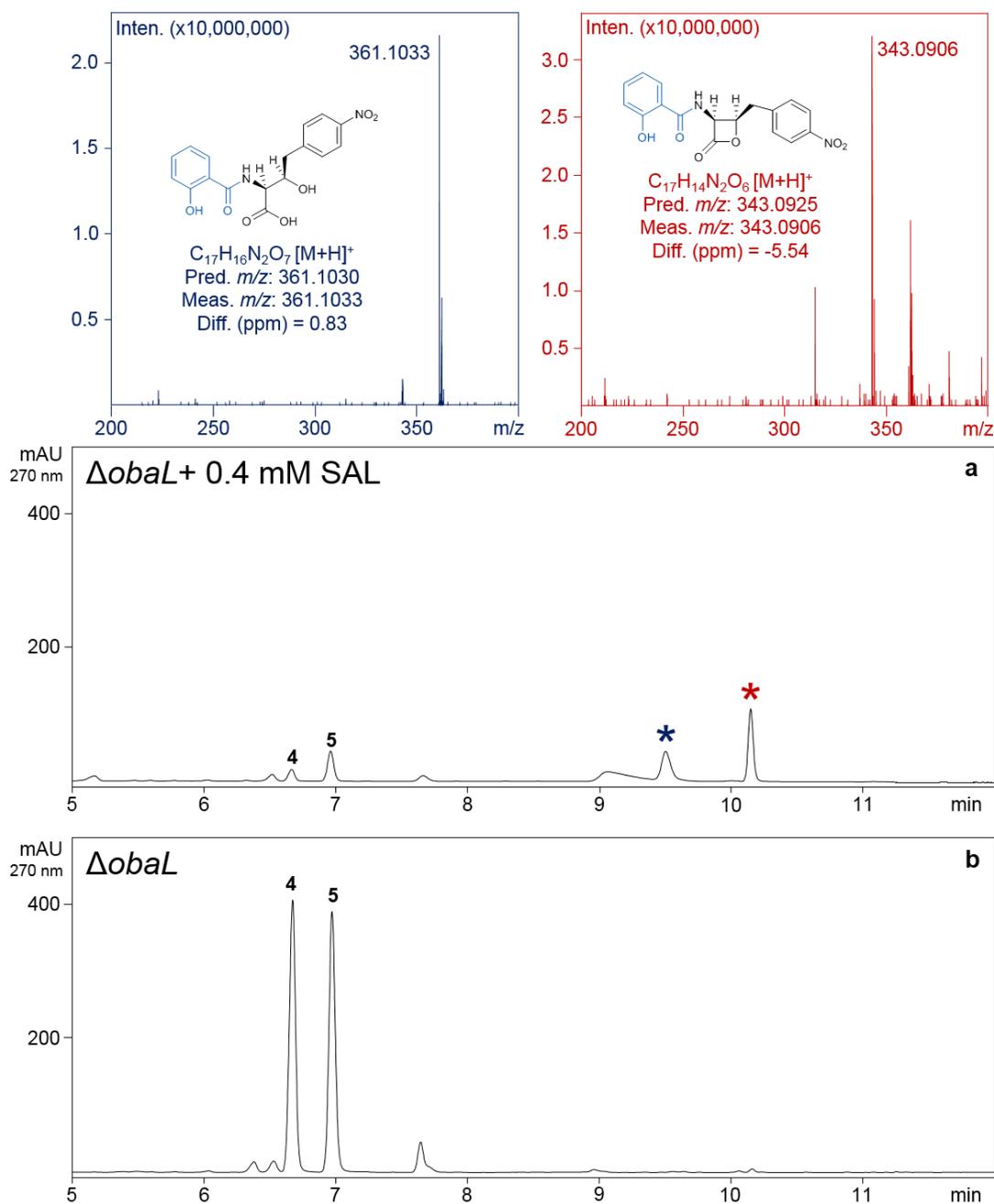


Figure 8.14. HPLC and MS data for Δ obaL fed 0.4 mM salicylic acid (SAL). (a) UV chromatogram for a Δ obaL production culture fed SAL to a final concentration of 0.4 mM. Extracted mass spectra for peaks identified in the SAL-fed culture UV chromatogram with * are included above: Red = SAL congener of 1 (Calculated for $C_{17}H_{14}N_2O_6$ [M+H]⁺, Pred. m/z : 343.0925, Meas. m/z : 343.0906, Diff. (ppm) = -5.54); Blue = Hydrolysed form of the SAL congener of 1 (Calculated for $C_{17}H_{16}N_2O_7$ [M+H]⁺, Pred. m/z : 361.1030, Meas. m/z : 361.1033, Diff. (ppm) = 0.83). (b) UV chromatogram of a Δ obaL production culture fed DMSO as a negative control.

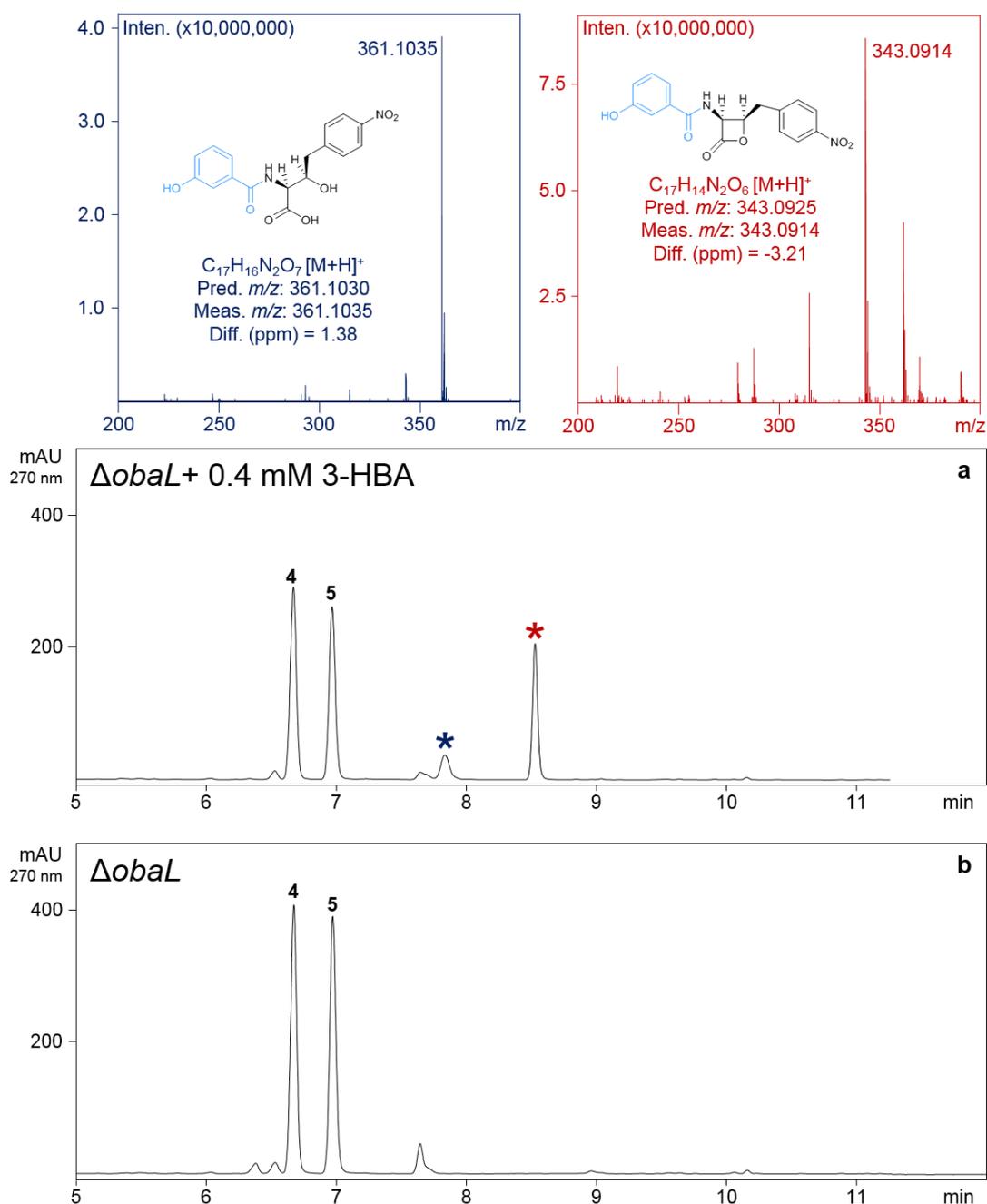


Figure 8.15. HPLC and MS data for $\Delta obaL$ fed 0.4 mM 3-hydroxybenzoic acid (3-HBA). (a) UV chromatogram for a $\Delta obaL$ production culture fed 3-HBA to a final concentration of 0.4 mM. Extracted mass spectra for peaks identified in the 3-HBA-fed culture UV chromatogram with * are included above: Red = 3-HBA congener of **1** (Calculated for $C_{17}H_{14}N_2O_6 [M+H]^+$, Pred. m/z : 343.0925, Meas. m/z : 343.0914, Diff. (ppm) = -3.21); Blue = Hydrolysed form of the 3-HBA congener of **1** (Calculated for $C_{17}H_{16}N_2O_7 [M+H]^+$, Pred. m/z : 361.1030, Meas. m/z : 361.1035, Diff. (ppm) = 1.38). (b) UV chromatogram of a $\Delta obaL$ production culture fed DMSO as a negative control.

8.7 Discussion

Access to genomic data for additional strains that contain putative *oba* BGCs allowed the identification of a putative resistance gene (*obaO*) for **1**, based on a minimal set of genes that are common to all clusters. *ObaO* encodes a ThrRS, the enzyme essential for loading L-threonine onto its cognate tRNA for protein synthesis. Although aaRSs are ubiquitous in nature, divergence in their sequence and structure has created the opportunity for organisms to be able to biosynthesise NPs that selectively target the aaRSs of potential competitors (Cochrane *et al.*, 2016). Self-resistance can also be achieved via the duplication and divergence of the target aaRS within the producer to create a resistant copy. This was the basis for the identification of *obaO*, which represents a second putative ThrRS encoded in the ATCC 39502 genome. This is also the case in the genomes of other bacteria that encode putative *oba* BGCs. Many other NP BGCs that encode inhibitors of aaRSs contain a resistant copy of the target aaRS, in addition to a primary copy encoded elsewhere in the genome. These include the mupirocin BGC which encodes a resistant isoleucyl-tRNA synthetase (El-Sayed *et al.*, 2003), the indolmycin BGC which encodes a resistant tryptophanyl-tRNA synthetase (Du *et al.*, 2015) and the BGC for borrelidin, also a ThrRS inhibitor, which encodes a resistant ThrRS (Olano *et al.*, 2004).

Disruption of the *obaO* gene in *P. fluorescens* ATCC 39502 led to the abolition of **1** production and loss of the characteristic purple phenotype in production cultures, indicating an intrinsic relationship between this locus and **1** production. Unfortunately, introduction of pJH10TS-*obaO* into the Δ *obaO* strain was unable to recover the WT production phenotype. This could be because overexpression of this gene disrupts the physiology of the cell (e.g. through L-threonine starvation), resulting in suppressors in which the gene is not expressed. A more likely explanation is that removal of *obaO* function leads to a suppressor mutation in a biosynthetic *oba* gene to abolish production and prevent self-harm. To circumvent this issue, in future work *obaO* will be introduced into the WT strain elsewhere in the genome, *before* subsequently disrupting the *oba* BGC cluster copy, thus preventing selection for biosynthesis mutants under **1**-producing conditions. This could be performed using pUC18T-mini-Tn7-Gm, a vector which integrates DNA sequences flanked by Tn7R and Tn7L coding sequences into Tn7 transposition sites. In the *Pseudomonas* chromosome, such an *att*Tn7 transposition site is located downstream of a highly-conserved glucosamine-6-phosphate synthetase (*glmS*) gene (Choi and Schweizer, 2006). A pTS1-based knockout can then be performed to target the endogenous *oba*

BGC copy, which should allow the possibility of non-polar effects as a result of *obaO* disruption to be ruled out.

Agar diffusion assays performed to identify the MIC of **1** against two *E. coli* strains showed that *obaO* increased the MIC by >60 fold for ATCC 25922, and >4000 fold for NR698, compared to empty vector controls. This indicated that *obaO* represents the **1** resistance gene and that **1** is an inhibitor of ThrRS. **1** was also shown to be active against *B. subtilis*, a Gram-positive bacterium, with an MIC of only 0.125 µg/mL. Further experiments are required to clone a codon-optimised copy of *obaO* in an appropriate expression vector into *B. subtilis*, in order to see whether it can also confer resistance in this bacterium. As a further control, the 'primary' metabolic ThrRS PCS from the ATCC 39502 strain could also be introduced in to bioassay strains to confirm that it does not confer resistance to **1**. In the future, bioassays can be extended to investigate **1** antibiotic activity against ESKAPE (Rice, 2008) and other clinically significant pathogens to establish the spectrum of its activity. To provide additional support for **1**'s role in ThrRS inhibition, future work will also include the use of reporter strains in which the promoters of genes known to be highly transcribed in response to different antimicrobial compounds have been fused to a luciferase reporter gene (Hutter *et al.*, 2004). When challenged with **1**, only strains with protein synthesis-associated gene promoters should express luciferase and display bioluminescence.

Alignment of amino acid sequences for putative **1**-'sensitive' and -'resistant' ThrRSs revealed that putative **1** resistant genes are substituted with a valine at a key ThrRS active site cysteine residue. Cysteine represents an ideal candidate nucleophilic residue for β -lactone ring attack, and would explain why putative resistance genes have evolved to encode a valine at this position to prevent reaction with **1**. To test this, variants of *E. coli* ThrRS were purified and incubated with **1** before HR-LCMS analysis to determine whether mass shifts corresponding to **1**-binding could be detected. Unfortunately, in both cases, a mass shift of ~358 Da was observed indicating that both variants bound **1**, and in neither case was a cysteine residue found to be modified. Future work will be to perform the same experiment with purified *obaO* and a V481C variant to identify whether the result observed for *E. coli* ThrRS was due to of technical issues or whether **1** might possess an alternative mode of action for ThrRS inhibition. The other ThrRS encoded in the ATCC 39502 genome will similarly be investigated (C479V substitution). Constructs in which cysteine and valine residues have been switched and vice versa will also be introduced into

bioassay strains to determine the significance of that residue for antibiotic activity. Samples of purified **1** have been sent to Prof. Christopher Francklyn's group at the University of Vermont, who specialise in the biochemistry of aaRSs, to perform kinetic analyses to further verify whether **1** is a true ThrRS-binder.

Not only is **1** prone to hydrolysis under aqueous conditions at neutral and alkaline pH, but it was also found to be particularly reactive with other chemicals in solution. During the course of experimental optimisation for **1** ThrRS-binding assays, it was found that **1** readily reacts with glycerol, in addition to a number of buffers including Tris and HEPES. Fortunately, MES buffer did not react with **1** and hydrolysis could be minimised at pH 6. This lability could explain several of the results observed in preliminary mode of action experiments; microtitre plate-based bioassays in MH broth might have failed if **1** reacts with any of the constituents in this medium for example. Instability of the molecule would also explain why relatively little experimental data on **1** activity are available in the literature. β -lactam rings are seemingly more prevalent among NPs compared to β -lactones, which are relatively uncommon, and this observation could also be a reflection in nature of issues associated β -lactone instability. The major force driving the reactivity in both of these ring-strained systems is the high relief on ring strain following opening (95-130 kJ/mol – Roux *et al.*, 1995; Kim *et al.*, 2002), however, the amide bond of the azetidine-2-one ring (β -lactams) is more stable to hydrolysis than the oxetan-2-one (β -lactone) ester bond. Whilst this causes the β -lactam ring to be less reactive, it provides significant advantage in minimising unspecific reactivity that would hamper the utility of these molecules in nature, as exemplified by experimental difficulties encountered with **1**.

In addition to the β -lactone ring, **1** comprises a catechol moiety which is often found in siderophore NPs and is responsible for metal-ion chelation. ThrRSs require Zn^{2+} as a cofactor for activity, raising the possibility that the **1** catechol might also be important for mode of action through cofactor-binding. This raised two possible modes of action based on both β -lactone ring and catechol moieties: 1) The binding of the β -lactone to the ThrRS active site cysteine residue is reversible, but is stabilised by catechol binding to the Zn^{2+} cofactor which acts to keep the β -lactone in close proximity to cysteine for subsequent reactions; and/or 2) Initial coordination of the Zn^{2+} cofactor is necessary to orient the β -lactone into the necessary position to facilitate attack of the ring by the active site cysteine residue. If coordination of the Zn^{2+} ThrRS cofactor is indeed necessary for **1**-mediated inhibition, this would explain why the specific modification of the cysteine residue could not be detected following

tryptic digestion in biochemical assays with *E. coli* ThrRS, as **1**-Zn²⁺ binding would be disrupted. However, it does not explain why **1**-binding was detected in both WT and C480V ThrRS variant protein assays prior to tryptic digestion.

Ability of **1** to chelate metal ions was assessed in preliminary CAS assay experiments, which indicated that **1** does indeed chelate Fe³⁺, as a colour change from blue to pink was observed, indicative of CAS-Fe³⁺ complex disruption by competing **1** to release CAS dye. However, potent ferric iron chelators cause the complete dissociation of the CAS-Fe³⁺ complex resulting in a distinct orange colour. In order to quantify **1** iron-binding, this assay will be repeated in future experiments over a range of concentrations and resulting absorbance spectra will be collected to quantify iron chelation by calculating EC₅₀ values for the decomplexation of CAS-Fe³⁺. A similar metal-binding competition assay was performed to assess the ability of **1** to outcompete PAR for Zn²⁺-binding. Whilst the characteristic Zn(PAR)₂ complex absorbance peak at 493 nm clearly decreased with increasing **1** concentrations, no concomitant increase in absorbance at ~410 nm reflecting dissociated PAR was observed. One possibility is that **1** could react with PAR independently of Zn²⁺ and that is why an increase in absorbance at ~410 nm is not observed when **1** is introduced. The basic pyridine moiety of PAR could aid hydrolysis of **1** and the hydroxyl groups of the resorcinol moiety are also likely to be relatively nucleophilic, and attack of the β-lactone ring could also be an issue. For both metal-binding assays, LCMS analyses are required to determine whether **1** reacts with any reaction components other than the metal ions. This would also allow determination of the ratio of **1** to each metal ion in complex, permitting quantification of metal ion-binding through binding/dissociation constant calculations.

An alternative possibility is that the catechol moiety is not at all involved in directing **1** ThrRS-binding, and instead, could play a role in uptake. Metal ions are often limiting in natural environments and many microorganisms biosynthesise siderophores to mediate ion-uptake (Wandersman and Delepelaire, 2004). The notable antibiotic activity demonstrated by **1** makes it unlikely that it functions purely as a siderophore, however, its siderophore-like properties may provide **1** entry into target cells. Many microorganisms have evolved the ability to utilise non-native siderophores (xenosiderophores), conferring them with a special fitness advantage by placing the metabolic burden of biosynthesis and secretion on others. This phenomenon is well documented in many key bacterial species (Jin *et al.*, 2006; Cornelis, 2010; Hammer and Skaar, 2011; Miethke *et al.*, 2013; Endicott *et al.*, 2017), and xenosiderophores

have also been successfully applied to promote the growth of uncultivated bacteria (D'Onofrio *et al.*, 2010). As a counter-evolutionary strategy, some microorganisms have evolved the ability to synthesise antibiotic-siderophore NP conjugates (sideromycins) that are taken up via siderophore transport proteins to circumvent issues of cell membrane permeability and kill competitors (Klahn and Bröstrup, 2017). Albomycins, for example, are composed of the natural antibiotic SB-217452 (Stefanska *et al.*, 2000), a seryl-tRNA synthetase inhibitor, and a hydroxymate-type ferrichrome moiety, known to be utilised as a xenosiderophore by *E. coli* (Hartmann *et al.*, 1979; Braun *et al.*, 1994; Clarke *et al.*, 2002), *P. aeruginosa* (Jurkevitch *et al.*, 1994; Hannauer *et al.*, 2010), *Streptococcus pneumoniae* (Pramanik and Braun, 2006) and *Salmonella* spp. (Ernst *et al.*, 1978; Braun *et al.*, 1983). Other natural examples include the salmycins (Vertesy *et al.*, 1995), ferrimycins (Bickel *et al.*, 1965), danomycins (Tsukiura *et al.*, 1964) and class IIb microcins (Duquesne *et al.*, 2007). **1** may then represent another example, in which the catechol moiety facilitates uptake into the target cell and the β -lactone mediates cell death. It is also possible that **1** export in ATCC 39502 is mediated by siderophore transporters, rather than by the RDD-domain protein encoded by *orf7* as suggested in Chapter 3, particularly given that no RDD-family proteins have been documented in the NP literature.

Future work in determining the mechanism of action for **1** will be aided by the mutasynthetic generation of four novel congeners of **1**. In each of these, the catechol has been substituted with different benzyl carboxylic acid moieties which should have reduced metal-binding abilities. This can be confirmed using CAS and PAR assays. If appropriate conditions for **1**-binding biochemical assay experiments can be established, congeners should be able to determine whether the catechol is necessary for ThrRS interaction, and thus distinguish between possible roles in uptake or ThrRS Zn²⁺ cofactor-binding for the **1** catechol moiety. Congeners of **1** can also be applied in *in vivo* experiments to determine how alterations in this moiety affect antibiotic activity. Production of **1** congeners will be scaled up so that sufficient material for each can be purified for NMR analysis and conformation of structure, and for use in biochemical and bioassay experiments. Whilst access to congeners of **1** allow a far more comprehensive study of the significance of the catechol moiety of **1**, a crystal structure for the **1**-ThrRS complex would permit the absolute characterisation of **1** ThrRS-binding mechanism.

A key phenotypic characteristic of **1** production cultures is the generation of a purple colouration (Figures 4.3 and 8.4). During *in silico* analysis of the ATCC 39502

genome, no BGCs for known potential candidate pigmented NPs, such as violacein (Lichstein and Vandesand, 1945) or prodiginines (e.g. prodigiosin, which gives *Serratia marcescens* its characteristic violet/red colour (Williamson *et al.*, 2006)), were identified. The colour is lost or at least diminished in all **1** biosynthetic deletion strains (Figure 4.3), indicating that it is either a direct or indirect result of **1** production. The observation that purified **1** alone is in fact pale yellow colour, and that **1** congener production cultures do not retain the purple phenotype (Figure 8.11), suggested that colour might be the result of a **1**-metal ion complex chromophore. Further work is required to elucidate the true origins of this intriguing production phenotype and will involve simple experiments incubating physiological concentrations of **1** with a range of metal ions in various concentrations to see if any colour develops.

The significance of the aromatic nitro group of **1** has not been explored at all in this work. Nitro groups are rare moieties among NPs (Winkler and Hertweck, 2007; Parry *et al.*, 2011), but have been shown to be critical to the bioactivity of several NPs including thaxtomin (King *et al.*, 2001; Scheible *et al.*, 2003), a phytotoxin that inhibits cellulose synthesis, and pyrrolnitrin, an antifungal inhibitor of electron transport (Lambowitz and Slayman, 1972; van Pée and Ligon, 2000). Currently, the best characterised involvement of a nitro group in NP bioactivity is from a structure of chloramphenicol in complex with the 70S ribosome from *Thermus thermophilus* (Bulkley *et al.*, 2010). In this structure, the chloramphenicol 4-nitrobenzyl group participates in a π -stacking interaction with C2452 of the ribosomal 23S RNA. Given that our data suggest ObaC operates at an earlier stage during the biosynthesis of **1** and that the nitro group appears to be significant for ObaH substrate recognition, it seems likely that nitro group incorporation has been specifically selected for. The strong electronegativity of the nitro group can delocalise the π -electrons of the benzene ring, providing a charge to the molecule which could be important for directing interactions with the ThrRS target (Ju and Parales, 2010). Unfortunately, the amino- congener of **1** does not accumulate in the Δ obaC mutant, and so cannot be compared to **1** in bioassay experiments. However, given that Schaffer *et al.*, (2017) demonstrated **1** assembly *in vitro* by incubation of **2** and **6** with ObaI, ObaK and Sfp, a mutasynthesis approach could be employed using (2S)-amino-(3R)-hydroxy-4-phenylbutanoate (prepared enzymatically using His₆ObaG) in place of **2** to generate the corresponding congener of **1**. This would only work provided the A₂ domain (and downstream-acting domains) is sufficiently promiscuous and will accept these alternate substrates. The hydroxylamine-trapping assay can be used to determine whether this is a possibility. Alternatively, **1** congeners substituted at the 4-nitro

position can be prepared chemically using published total synthesis methods for **1** (Lowe *et al.*, 1992). As mentioned previously, a crystal structure of the **1**-ThrRS complex will permit elucidation of the precise interactions that mediate **1**-binding.

Whilst the results presented here represent very preliminary data, they reveal that **1** has very potent antibiotic activity against both Gram-positive and Gram-negative bacteria. Furthermore, they suggest that the β -lactone ring, with possible involvement of the catechol moiety, may represent a novel mechanism of inhibition for ThrRSs, and future work will be concerned with fully characterising the **1** mode of action, which will be aided by the generation of novel congeners of **1**. If **1** does indeed have a completely novel mode of action, it would warrant further experiments to explore its potential for therapeutic use and improve its drug-like properties.

Chapter 9:

Discussion

9. Discussion

9.1 Summary of results

In this project, mutational analyses were performed to dissect the biosynthetic pathway for obafluorin (**1**) using custom made suicide and complementation vectors. The precursors 2,3-DHBA (**6**) and 4-NPP (**8**) (via 4-APP (**7**)) were shown to be the products of gene sub-clusters (*obaJLN* and *oba(C)DEF*, respectively) in the pathway. **8** is an intermediate in the biosynthesis of AHNB (**2**), and both **2** and **6** are activated and condensed by a dimodular NRPS (Obal) in conjunction with the discrete ArCP (ObaK) to yield **1**. Failure of **2** to complement biosynthetic mutants precluded the possibility of characterising the biosynthesis of **2** from **8** *in vivo*, necessitating further biochemical experiments. The ThDP-dependent enzyme ObaH was found to act as an **8** decarboxylase, responsible for generating the intermediate 4-NPA (**11**), which is the substrate of ObaG. This disproved the previous hypothesis that **2** is the direct product of a ThDP-dependent mechanism (Herbert and Knaggs, 1990; 1992b). Instead, ObaG, a PLP-dependent L-TTA, was shown to be responsible for the biosynthesis of **2** as a single stereoisomer in respectable yields from L-threonine and **11**, with acetaldehyde as a by-product. ObaG represents a third subfamily of the L-TTAs, adding an additional member to join the FTase and LipK-like subfamilies.

L-TTAs, SHMTs and L-TAs all catalyse the cleavage of β -OH- α -AA substrates to produce glycine and an aldehyde by-product, and a phylogenetic analysis was performed to gain insight into the evolutionary history of these fold-type I PLP-dependent enzymes. All three enzyme families were determined to originate from an ancestor that probably used L-serine as a substrate and two separate substrate-switching events led to L-TA and L-TTA lineages that use L-threonine. The phylogeny allowed the identification of a new candidate L-TTA family, homologues of which are involved in the biosynthesis of β -OH- α -AA intermediates in several clavum metabolite pathways. Phylogenetics may represent a powerful approach for the expansion of the L-TTA enzyme class, and may additionally allow the identification of lineages of PLP-dependent enzymes that perform further unusual chemistries in NP biosynthetic pathways.

Biochemical experiments were also performed on the dimodular NRPS Obal, which is responsible for **1** assembly from **2** and **6**. The offloading of the newly assembled *pseudo*-dipeptide occurs with concomitant β -lactone ring formation. A hydroxylamine-trapping assay was performed to probe the substrate specificity of the Obal A

domains, and identified **2** as the preferred substrate, although significant signals were recorded for L-serine, L-tryptophan, benzoic acid (BA) and 3-hydroxybenzoic acid (3-HBA). This indicated some degree of substrate promiscuity in both A domains. The Oba1 TE domain contains an active site cysteine residue that was shown by others (Schaffer *et al.*, 2017) to be critical in creating thermodynamically favourable conditions for cyclisation and formation of the strained β -lactone ring in **1** to occur. With this data, the biosynthetic pathway to **1** from chorismate has been effectively delineated.

Finally, *in silico* analysis of putative *oba* BGCs encoded in the genomes of bacteria uploaded to GenBank during the course of this project permitted the identification of *obaO* as the putative **1** resistance gene. It encodes a ThrRS, essential for ligating L-threonine to its cognate tRNA during protein biosynthesis. Its role was confirmed by mutational analysis and agar diffusion assays, which showed that **1** has potent antibiotic activity against both Gram-positive and -negative bacteria, and that *obaO* is sufficient to confer high-level resistance upon sensitive strains. Alignments of ObaO homologues compared with those of putative primary metabolic ThrRSs revealed a C480V (*E. coli* numbering) substitution in 'resistant' copies. However, *in vitro* binding assays with WT and C480V copies of *E. coli* ThrRS have so far failed to shed light on the possibility that this is the nucleophilic residue that attacks the β -lactone ring to bind **1**. Preliminary experiments were also performed to assess the metal-ion-binding capacity of **1**, given that ThrRSs require a Zn^{2+} cofactor for catalytic activity and **1** comprises a catechol moiety synonymous with a role in Fe^{3+} chelation in NP siderophores. CAS and PAR assays indicated that **1** binds both Fe^{3+} and Zn^{2+} , but further work is required to dissect the mechanism of **1** ThrRS-binding. A mutasynthetic approach was applied by feeding alternate benzyl carboxylic acids to the Δ *obaL* strain and resulted in the generation of at least three novel congeners of **1** in which one or both catechol hydroxyl groups are no longer present. These congeners will form the basis of further bioassay and metal-binding experiments to determine the significance of the catechol moiety in the antibiotic activity of **1**.

9.2 Future directions

9.2.1 Regulation of the biosynthesis of obafluorin (**1**)

Although the potential quorum-dependent regulatory roles of *obaA* and *obaB* (*luxI* and *luxR* homologues respectively) were discussed in Chapter 3, no experimental work was performed in this project to characterise their respective functions during the biosynthesis of **1**. *luxR* homologues have been implicated in the regulation of

many pseudomonad NP biosynthetic pathways (Pierson III *et al.*, 1995; El-Sayed *et al.*, 2001; de Bruijn *et al.*, 2008; Dubern *et al.*, 2008), where, presumably, they ensure that a compound is only made when it can be produced in biologically significant concentrations. Logical initial experiments to explore the regulation of the biosynthesis of **1** would be to first define a production time course to identify when peak **1** production occurs and at what cell density. Experiments can then be performed to see whether peak production can be shifted to an earlier or later time point by introduction or removal of cell density. It would be interesting to see whether the purple colour that coincides with **1** production also develops earlier or later, as it remains unclear whether this colour is a direct or indirect consequence of the biosynthesis of **1**. Additionally, disruption and/or overexpression of either *obaA* and/or *obaB* should result in either an increase or decrease in **1** titres relative to the WT, which would indicate whether these regulators are acting as repressors or activators. In addition to genetic complementation, co-culture experiments with Δ *obaB* might also be an option to recover **1** production in an Δ *obaA* strain, as the autoinducer synthesised by ObaA in the Δ *obaB* strain cells should diffuse in liquid culture, eliciting a response from ObaB in the Δ *obaA* cells, thus switching **1** production on (or off). Using concentrated extracts from spent Δ *obaB* (or biosynthetic mutant) culture medium should produce a similar result in the Δ *obaA* strain.

In silico analysis of the *oba* BGC should allow the identification of putative Lux(Oba)-boxes, 20 bp inverted repeat sequences located ~40 bp upstream of transcriptional start sites, to which LuxR homologues bind to mediate gene regulation (Devine *et al.*, 1989; Qin *et al.*, 2007). Identified targets can be identified via several methods including reporter fusions and reverse transcriptional analysis to determine which genes are genuinely regulated by ObaB. The ultimate experiment to identify all the genes in the ObaB regulon would be to perform Chromatin Immunoprecipitation Sequencing (ChIP-seq) which, coupled with transcriptional organisation mapping, would identify regions of DNA to which ObaB is bound under particular growth conditions. With ObaB targets in hand, the *N*-acylhomoserinelactone molecule to which ObaB responds can be identified using a bioautography approach, as used by Thomas and co-workers in the study of quorum-sensing mediated regulation in mupirocin biosynthesis (El-Sayed *et al.*, 2001).

9.2.2 L-TTA directed evolution

L-TTAs represent exciting tools for synthetic applications in the generation of β -OH- α -AAs, far surpassing the aldol reaction catalysed by engineered SHMTs and L-TAs

in terms of both product yield and stereoselectivity. Whilst ObaG was shown to demonstrate some aldehyde substrate promiscuity (these enzymes are highly selective for their L-threonine amino acid donor), the only other characterised L-TTA family members, LipK and FTase, have not been explored for the generation of novel β -OH- α -AAs. This might be one direction for further study provided appropriate alternate aldehyde substrates can be prepared. Rational engineering by directed evolution of these enzymes, after solving their crystal structures, represents another possible strategy that might expand our access to new β -OH- α -AAs.

Compared to their primary metabolic counterparts, specialised metabolic enzymes tend to be significantly less efficient, as illustrated by the relatively high K_m values observed for ObaG and LipK. High efficiency is expected for central metabolic enzymes as the reactions they catalyse are intrinsically linked to growth rate and so strong selection pressures have been acting to optimise their catalytic activity over many millions of years (Bar-Even *et al.*, 2011). This would explain why SHMTs and L-TAs have proven to be recalcitrant in efforts to engineer them for β -OH- α -AA biosynthesis (Fesko, 2016). In specialised metabolism, there appears to be significant selection for the maintenance of chemical diversity (Firn and Jones, 2000), and broader substrate specificity is selected for at the expense of catalytic efficiency, which also serves to avoid competition with essential primary metabolic pathways (Bar-Even and Tawfik, 2013). As a result, L-TTAs may be more amenable to manipulation for altered substrate specificity.

In nature, specialised metabolic biosynthetic genes are thought to have evolved via a number of different routes including HGT, homologous recombination, gene fusion and gene duplication followed by divergence (Eisenbeis and Hocker, 2010). In the latter case, a duplicated primary metabolic enzyme is proposed to first diverge to become a substrate generalist, before evolving into a specialist enzyme optimised for the use of a new specific substrate in the context of NP biosynthesis (Firn and Jones, 2000). A similar logic would be employed in a directed evolution approach with L-TTA enzymes, using rounds of gene-sequence level diversification followed by expression, screening and selection for alterations that lead to better acceptance of a desired alternate substrate (Denard *et al.*, 2015). Once a new substrate is accepted, similar approaches can be used to select for improved yield and stereoselectivity to make the enzyme suitable for use as a biocatalyst. Optimisation of protein expression or thermostability for industrial use can also subsequently be considered. This work would be facilitated by a greater understanding of the reaction mechanism employed

by these enzymes and structural data for each of the L-TTAs, especially if they can be crystallised with their substrates bound. As mentioned in Chapter 6, this would allow the identification of key residues for modification in the active site that are necessary for substrate interaction and/or for creating the three-dimensional space necessary to accommodate new substrates.

9.2.3 Strategies for future genome mining efforts

One of the key broader questions this project hoped to address was how can we mine the wealth of available genomic data in a way that selects for discovering novel chemistry? Accessing new, diverse chemical structures is paramount to identifying NPs that interact with new biological targets, and will be critical in tackling the issue of multidrug resistant pathogenic bacteria. Initially, it was hoped that a ThDP-dependent enzyme and transaminase, as previously hypothesised to be responsible for the biosynthesis of **2** (Figure 3.1), could be used as a 'fingerprint' to query the genomic sequence databases to identify BGCs that biosynthesise unusual β -OH- α -AAs or α -OH- β -AAs. In this project however, **2** was proven to be the product of a PLP-dependent L-TTA reaction.

This finding does not prevent the development of alternate strategies to search for novel chemistries, as highlighted by the use of phylogenetics to identify further potential L-TTAs among clavam biosynthetic pathways. In collaboration with Dr. Govind Chandra, the phylogenetic study will be expanded to include more putative SHMTs, L-TAs and related PLP-dependent enzymes (e.g. methylSHMTs) available in the GenBank database. This may allow us to both identify further L-TTAs that have been incorrectly annotated, and identify additional clades of PLP-enzymes that have diverged from primary metabolic enzymes to fulfil roles in NP biosynthetic pathways. The clustering of PacT homologues, thought to be responsible for the biosynthesis of DABA in four different biosynthetic pathways (Zhang *et al.*, 2010; Kaysser *et al.*, 2011; Li *et al.*, 2013; Jiang *et al.*, 2015), indicates that this is possible. The PLP-cofactor is well known for its catalytic versatility (Contestabile *et al.*, 2001) and so these enzymes represent ideal candidates for divergence into new functions in specialised metabolism. Indeed, many examples of PLP-dependent enzymes and domains have been reported in the NPs literature (highlighted in Chapter 6). Phylogenetics may therefore allow us to unlock a broader range of chemistries for PLP-dependent enzymes that have evolved to operate in NP biosynthetic contexts.

Given that the product of all L-TTAs is an amino acid, more refined genome mining attempts could be envisaged using characterised family members in coupled queries with NRPS domains/modules, as these represent the most common mechanism of amino acid assembly among NPs. ATP-grasp ligases could be similarly applied as these are also reported to introduce amino acids into NPs (Dawlaty *et al.*, 2010; Noike *et al.*, 2015; Ooya *et al.*, 2015). For example, Lig is an ATP-grasp fold family protein encoded in the valclavam BGC that is proposed to ligate valine and a 9-hydroxycavam intermediate (**Z** in Figure 6.2 - Nobary and Jensen, 2012); this work suggests **Z** is the product of an L-TTA-catalysed reaction. However, biasing a query based on amino acid assembly would exclude the discovery of L-TTAs such as FTase (Murphy *et al.*, 2001) or OrfA (Zelyas *et al.*, 2008) in which the β -OH- α -AA alone forms the entire scaffold for the final product. Not only would crystal structures for L-TTA subfamily members aid in guiding directed evolution efforts to engineer these enzymes, but they would also facilitate more directed searches for additional L-TTA family members. The characterisation of key catalytic residues necessary for aldehyde substrate interaction in these enzymes would provide an L-TTA 'signature' for genome-mining efforts.

ThDP-dependent enzymes also catalyse an incredibly diverse range of biotransformations in nature, including cleavage and formation of C-C, C-O, C-N and C-S bonds (Pohl *et al.*, 2004; Duggleby, 2006). Diversity in substrate specificity and catalytic activity have been the result of the shuffling, rearrangement and fusion of domains, mutations and gene duplications, which has created a great breadth of ThDP-dependent enzymes (Vogel and Pleiss, 2014). Like PLP, ThDP is also able to catalyse reactions independently in solution, albeit at a much slower rates, indicating that early ThDP-dependent enzymes might have exhibited broad substrate specificities (Kluger and Tittmann, 2008). ThDP-dependent enzymes all share a common reaction mechanism in which a ThDP ylide is formed by protein-assisted deprotonation at its C-2 atom, which subsequently acts as a nucleophile adding to the α -carbon of a 2-keto substrate (Jordan, 2003). Pyrophosphate (PP) and pyrimidine (PYR) domains are always present and are necessary for binding and activating the cofactor. Perhaps unsurprisingly, these versatile catalysts have also diverged into NP biosynthetic contexts to utilise new unusual substrates (e.g. ObaH) and perform novel chemistries (Chen *et al.*, 1998; Balskus and Walsh, 2008; Peng *et al.*, 2012; Proschak *et al.*, 2014; Su *et al.*, 2016). However, the only phylogenetic study of these enzymes to date included just 17 different ThDP-dependent enzyme amino acid sequences (Costelloe *et al.*, 2008), warranting a more expansive study to

further understand their evolutionary history. This might also allow the identification of lineages involved in NP biosynthesis.

Phylogenetics might also be useful tool for the identification of further BGCs encoding β -lactone and β -lactam NPs. This would be extremely desirable given their reputations for being potent specific inhibitors of important biological targets. As mentioned earlier, Schaffer *et al.* (2017) demonstrated that β -lactone formation by Obal is catalysed by an active site cysteine, rather than a canonical serine, residue. This switch makes the formation of the strained ring thermodynamically favourable. Schaffer *et al.* also showed that the Obal TE domain and its homologues from putative **1** BGCs identified in Chapter 8, form their own discrete lineage in a phylogeny of class I and II TEs from other NP biosynthetic pathways. Given the apparent necessity of the cysteine residue for cyclisation (substitution to serine resulted in the hydrolysis of a linear dipeptide product (**10**) – Schaffer *et al.*, 2017), a phylogeny of all annotated TEs might allow the expansion of this clade with more distantly related domains that introduce β -lactone rings into different NPs. The recent example of sulfazecin biosynthesis in which the TE domain from SulM with an active site cysteine residue catalyses β -lactam ring formation (Li *et al.*, 2017), creates the potential to also identify novel β -lactam NP BGCs.

Gene duplication has played a key role in the evolution of NP biosynthetic pathways as additional gene copies have evolved to fulfil new functions in the context of specialised metabolism (Fischbach *et al.*, 2008). In the biosynthetic pathway to **1**, the L-TTA ObaG, and the ThrRS ObaO studied in this work are excellent examples of duplicated primary metabolic enzymes that have evolved to perform novel chemistry and resistance functions, respectively. During the evolution of bacterial genomes, if a duplicated gene does not provide some selective advantage, it will accumulate mutations, and its removal will be selected for due to the energy expenditure and lowered replication rates associated with the maintenance of additional DNA sequence (Sela *et al.*, 2016). The large effective population sizes of prokaryotes, in addition to short generation times, enables them to maintain very compact genomes by this process of selection (Lynch, 2006). Consequently, bacterial genomes will often only encode a single copy of each essential housekeeping gene, which has allowed their identification by approaches such as transposon mutagenesis (Simon *et al.*, 1983). This phenomenon also formed part of the basis for identification of *obaO* as the putative **1** resistance gene, which exists as the only additional copy to a putative housekeeping ThrRS encoded in the ATCC 39502 genome. This is the case in all

strains encoding a putative *oba* BGC, whereas model *P. fluorescens* strains F113, Pf0-1 and SBW25, which do not encode a putative *oba* cluster, only encode only a single *thrRS* gene. These observations raise the possibility that genetic redundancy could be utilised during genome mining approaches, as a marker of NP BGC evolution.

The essential housekeeping genes for many clinically relevant pathogenic bacteria have already been determined (Ackerley *et al.*, 2001; Geoffroy *et al.*, 2003; Garsin *et al.*, 2004; Liberati *et al.*, 2006). Essential genes for which a NP inhibitor is currently not known could be selected as queries, and a search of the genome sequence databases could be performed to identify strains which possess multiple copies of the selected essential gene target. Duplication of essential enzymes *is* known among bacteria (the *B. subtilis* genome encodes two ThrRSs (Putzer *et al.*, 1990)) so a condition could be applied to the query that one of the multiple copies of the target gene in a given strain must be within a certain distance of a known NP biosynthetic gene (e.g. NRPS/PKS). This caveat capitalises on the natural tendency for NP biosynthetic genes to be clustered together (Fischbach and Walsh, 2006; Osbourn, 2010). Hits would then have to be analysed manually to identify whether they do occur in plausible BGCs or not. False positives can be minimised by increasing the copy number threshold of a target gene, decreasing the maximum distance a target must be from a NP biosynthetic locus and by omitting species from the search in which redundancy of essential genes is particularly common.

This redundancy-based approach is limited in that it will not identify NP BGCs which rely on efflux or target modification as resistance mechanisms, will only identify BGCs comprising genes known previously to be involved in NP biosynthesis (NRPSs etc.), and can only be used with complete genomic data sets. However, the possibility of identifying BGCs for NPs with pre-determined targets could be very powerful in tackling the issue of multidrug resistant bacterial pathogens. In the first instance, redundancy-based genome mining could be performed and refined with actinomycete bacterial genomes as these are known to encode among the highest numbers of BGCs (Bérdy, 2012). Given that NPs that target aaRSs commonly employ a duplicated copy of the target enzyme as a resistance mechanism (Cochrane *et al.*, 2016), they would make ideal initial queries to validate this genome mining strategy.

Whilst developing this idea with Dr. Govind Chandra, a new web server called the Antibiotic Resistant Target Seeker (ARTS – Alanjary *et al.*, 2017) became available

online which detects known and putative resistance genes present within a submitted DNA sequence query based on duplication, localisation within a BGC and evidence of HGT. This web tool was able to recognise duplications of genes encoding putative SHMTs, ThrRSs and GCS genes in the ATCC 39502 genome, however it did not identify any of these as belonging to a putative BGC. Nevertheless, this tool shows great potential and indicates that a refined redundancy-based approach could work. The ARTS server has the added benefit of not being limited to type of resistance mechanism, and is able to identify putative efflux resistance mechanisms and NP modifying enzymes, in addition to resistant NP target enzymes.

Similar reasoning and database queries could conceivably be used to identify essential enzymes that have been acquired by NP biosynthetic pathways following duplication and have evolved to perform novel chemistry. This would also require queries to be coupled with known biosynthetic genes to reduce the numbers of hits located outside of NP BGCs. A recent phylogenomics-based approach designed to search for the expansion of central metabolic families among 230 actinobacteria strains was successful in identifying biosynthetic pathways to arseno-organic metabolites in *S. coelicolor* and *S. lividans* (Cruz-Morales *et al.*, 2016). This work further validates a redundancy-based genome mining strategy, but with the added benefit of not limiting hits to BGCs that encode members of characterised NP families. However, the presence of a duplicated essential gene in a putative BGC is by no means a guarantee of novel chemistry and there is no way of knowing this until at least the NP structure has been characterised.

9.2.4 Evolution of the *oba* BGC

The ability to rapidly generate structural variety seems to be an especially important property of NP biosynthetic pathways and is reflected in the fact that many are known to be responsible for the production of multiple compounds that are variations on the same core structure. As an example, work in the Wilkinson group on the formicamycins produced by *Streptomyces formicae* revealed that a single type II PKS BGC was responsible for the production of 13 natural detectable analogues of these pentacyclic PKs (Qin *et al.*, 2017). High-affinity, reversible, non-covalent interactions between a NP ligand and its protein target require the ligand to possess a highly specific configuration to interact with the specific 3D structure of its target, which is a rare property for any molecule to possess (Firn and Jones, 2000). This is reflected in the fact that different formicamycins vary greatly in their bioactivity against MRSA and VRE, and also in the failure of combinatorial chemistry and HTS in the late 1990s to

produce viable drug leads. However, the moderate, varied bioactivities exhibited by congeners provide sufficient selective advantage to favour the maintenance of chemical diversity in NP biosynthetic systems. In pathways for which the biosynthesis of a single product has been strongly selected for, BGCs remain sufficiently plastic that minor mutations or alterations can restore the ability to generate chemical diversity. This is illustrated by the tenellin biosynthetic pathway in the ascomycete fungus *Beauveria bassiana*, which was found to produce 18 new compounds following the silencing of individual genes by RNA interference or when epigenetic elicitors were introduced exogenously to production cultures (Yakasai *et al.*, 2011). Microbes inhabit highly dynamic environments and the ability to generate chemical diversity presumably allows them to adapt to new environmental conditions/competitors very rapidly.

The diversification of BGCs and the NPs they encode is facilitated by a number of different mechanisms, and evidence for several of these can be illustrated in the **1** biosynthetic pathway. One mechanism is via alterations to individual genes that cause them to accept new substrates or to perform new chemistries. This has been described in this work for ObaG, which appears to have evolved from an SHMT-like progenitor to accept L-threonine (instead of L-serine) as a substrate. Furthermore, it also catalyses a subsequent C-C bond-forming reaction with an aldehyde following L-threonine cleavage to generate an unusual β -OH- α -AA building block (**2**). The evolutionary radiation of lanthipeptides in marine cyanobacteria to produce diverse families of cyclic peptides represents an example in which mutations in precursor substrate genes have generated enormous chemical diversity (Donia *et al.*, 2006; Cubillos-Ruiz *et al.*, 2017, Gu and Schmidt, 2017). In larger multi-modular biosynthetic genes, such as PKSs and NRPSs, mutation of individual domains can either render them inactive, as is commonly observed in PKS processing domains (Fischbach *et al.*, 2008), or alter their substrate recognition. The Obal A₂ domain for example must have evolved from an A domain with specificity for one of the 20 canonical amino acids to accept its nonproteinogenic substrate **2**. Homologous recombination in these highly modular systems also represents a significant mechanism by which cluster and product variation is achieved (Fischbach *et al.*, 2008).

NP BGCs can also diversify through the acquisition or loss of entire genes, which can occur both intra- and intergenically (via HGT). As mentioned previously, non-functional genes are rapidly lost in bacterial genomes and this is no less true for NP

BGCs. Consequently, many are assembled from sub-clusters of genes which each encode their own small molecules that possess some degree of advantageous bioactivity, and so are maintained, available for incorporation into new biosynthetic contexts (Medema *et al.*, 2014b). The catechol moiety of **1** is encoded by the sub-cluster *obaJLN* and is an example of this phenomenon. **6** can act directly as a siderophore (López-Goñi *et al.*, 1992), albeit not as effectively as **6**-incorporating siderophores, the pathways of which might have evolved by recruiting the genes necessary to connect several **6** molecules together to augment Fe³⁺-binding capacity (Fischbach *et al.*, 2008). We propose that the *oba* BGC was formed by the merging of a sub-cluster of genes necessary for the biosynthesis of **6**, and a small cluster encoding a β -lactone NP. This would have been selected for because the **6**-derived catechol either mediated specific interaction with a new target (ThrRS via Zn²⁺ cofactor-binding) and/or improved **1** uptake into target cells via siderophore transporters. The putative *oba* BGC identified in several *Burkholderia* spp. (Figure 8.2) would represent an intermediate stage in the evolution of the BGC preceding the fusion of the A₁ domain necessary for **6** activation to Obal (possibly mediated by the embedded MbtH-like domain), but following the separation of an *entB*-like precursor to mobilise the cognate ArCP. We propose that a mobile ArCP is necessary to facilitate interaction with both the **6**-specific A domain and Obal homologue C domain. It is possible that the **2**-derived moiety was introduced in a similar fashion, following the merging of a progenitor NRPS cluster with the *obaCDEFGH* sub-cluster, replacing the previous proteinogenic amino acid incorporated into the β -lactone NP following evolution of the necessary A domain to accept **2** as a substrate. It is likely that *obaCDEFGH* operated together to biosynthesise **2** prior to their forming a sub-cluster, but clustering was selected for to coordinate the regulation of these genes and facilitate their co-inheritance by HGT. The **2**-derived moiety of **1** must also offer some selective advantage to have been maintained in all putative *oba* BGCs identified, possibly through mediating some other interaction in the ThrRS active site as mentioned previously. The nitro group of **1** could play an important role in this regard, as described in Chapter 8.

1 was initially discovered as part of efforts to discover novel β -lactam NPs using a β -lactamase induction assay (Wells *et al.*, 1982; 1984), during the course of which several other β -lactones were isolated and characterised from *Arthrobacter*, *Bacillus* and *Pseudomonas* isolates. These all differ from **1** in the composition of their C-2 and C-3 side chains (Figure 9.1a). In terms of BGC evolution, this suggests that an ancestor BGC may have comprised a monomodular NRPS minimally comprising

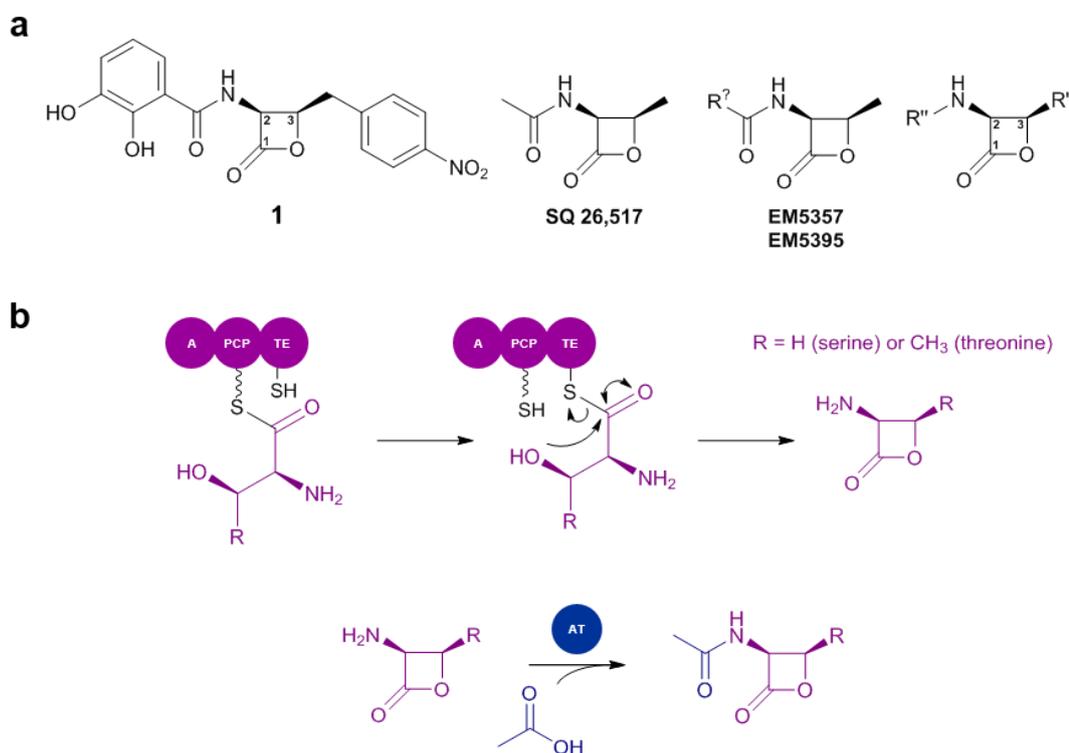


Figure 9.1 Evolutionary origins of the *oba* BGC. (a) β -lactones isolated by Wells *et al.* (1982; 1984) as part of screening efforts for β -lactam antibiotics. R² = an unspecified R group (Wells *et al.* 1982); R' and R'' are variable groups that could be altered using synthetic chemistry and/or biology approaches to develop novel β -lactones. (b) Hypothetical monomodular Obal ancestor NRPS. Minimally this comprises of A and PCP domains for the activation and tethering of either of the proteinogenic β -OH- α -AAs L-serine or L-threonine, and a TE domain with an active site cysteine for β -lactone ring formation. A hypothetical primary metabolic AT is proposed to introduce acetate to generate β -lactones like SQ 26,517.

three domains: an A domain specific for activation of a proteinogenic β -OH- α -AA substrate (L-serine or L-threonine), a PCP domain to tether it, and a cysteine active site residue-bearing TE domain responsible for β -lactone ring formation (Figure 9.1b). Chemical diversity has been generated in this system by the acquisition of genes necessary for biosynthesis of alternate acid substrates and the corresponding recruitment/evolution of A domains able to accept them. Two of the other β -lactones isolated by Sykes *et al.* comprise a methyl group as the C-3 side chain, indicating that the A domain in their hypothetical NRPSs utilise L-threonine as a substrate (Wells *et al.*, 1982). Initial *in silico* analysis of the Obal A₂ domain also identified L-threonine as its likely substrate, and it is conceivable that it evolved from being specific for L-threonine to **2** following the acquisition of *obaCDEFGH* to the BGC.

The C-2 side chain in several of the identified β -lactones appears to be derived from acetate, which is likely installed on the amino group by a background AT (SQ 26,517 - Figure 9.1a and c). Due to their nucleophilicity, primary amino groups are often modified to avoid non-specific reactions. This is exemplified by the desosamine moiety in methymycin; when the *N*-methyltransferase *desVI* (responsible for methylation of the desosamine amino group) was knocked out, only *N*-acylated intermediates were detected (Zhao *et al.*, 1998). If a C domain were also present on a monomodular NRPS ancestor, then the acetate group could alternatively be introduced via an acyl-S-ACP as observed in hybrid PKS/NRPS systems.

I favour this hypothesis of cluster evolution over one in which the Obal NRPS TE-domain active site serine has been mutated to a cysteine residue *after* amino acid substrates have been fixed, because this would likely require the simultaneous substitution of two bp to avoid an inactive intermediate. Perhaps more importantly, the inherent reactivity of the ring-strained β -lactone would mean that any product generated from a minimal monomodular NRPS described would have some degree of bioactivity to be selected for, maintained, and subsequently 'improved' upon, similar to the **6** example described previously. The same is not as certain for a linear dipeptide product, and the occurrence of other β -lactones that vary in the composition of the same two side chains supports the former hypothesis. A candidate progenitor BGC in which the NRPS TE comprises an active site serine has also not yet been identified.

An exciting future project would be to obtain the strains responsible for producing the other β -lactone molecules characterised by Sykes *et al.* (Wells *et al.*, 1982) and to

sequence their genomes to identify and characterise the BGCs responsible for producing them. This would allow the above hypotheses to be evaluated and perhaps revised, and might provide further understanding of how these different β -lactone biosynthetic pathways have evolved. Coincidentally, they would also permit further validation (or not) for the use of a TE-based phylogenetics approach to identify further β -lactone clusters. It would be particularly exciting if these molecules interact with different targets to **1**, as this might provide valuable insights into how different side chains have been selected for.

9.2.5 Synthetic biological manipulations of the *oba* BGC

A logical progression to the evolutionary work discussed would be to modify the C-2 and C-3 side chains of **1** using synthetic biology approaches. The primary focus of this work would be on engineering the specificity of the Obal (or analogous β -lactone synthesising NRPSs) A domains, as these will to a large extent determine the nature of the side chains, though the importance of downstream acting domains (C and TE) and the MbtH-like domain in substrate activation must also be considered (Meyer *et al.*, 2016). Mutasynthetic approaches have already been applied in this work to alter the C-2 side-chain and hydroxylamine-trapping assay results suggest that there is some degree of promiscuity in the A₂ domain as L-serine and L-tryptophan were both accepted as substrates. However, the latter of these does not possess the β -hydroxyl group necessary for β -lactone ring formation and a product without a C-3 side-chain (expected if L-serine used instead of **2**) was never detected in production cultures. This does not preclude the possibility that **1** congeners are subject to proteolytic degradation in the cell, and *in vitro* mutasynthetic experiments might allow the generation of this anticipated product in addition to others.

The modular nature of NRPSs has allowed for domains or even entire modules to be swapped to introduce new amino acids into the peptide backbone. A domain-swapping strategy was successfully employed to generate novel pyoverdine analogues with modified peptide backbones in *P. aeruginosa* PA01 by swapping either A domains or C-A didomains (Calcott *et al.*, 2014), for example. However, this strategy often encounters problems as even slight alterations to the modular NRPS structure can impact protein folding or the interactions between different domains, which is presumably why a previous attempt to alter the penultimate amino acid in pyoverdine failed (Ackerley and Lamont, 2004). This could prove detrimental in the Obal system depending on how both A domains interact with the MbtH-like domain and downstream C and TE domains.

More subtle alterations of A domain specificity can be achieved through site-directed mutagenesis of Stachelhaus motif residues to alter substrate specificity (Stachelhaus *et al.*, 1999). This approach has been applied to change A domain substrate specificity for both natural (Eppelmann *et al.*, 2002) and non-natural amino acids (Thirlway *et al.*, 2012) to generate novel NP analogues. However, binding pocket mutagenesis can be laborious and inappropriate for larger specificity changes and an alternate subdomain swapping strategy could also be considered. *In silico* analysis of the hormaomycin NRPS genes revealed that a natural recombination event between core subdomains comprising substrate determining active site residues likely occurred between A domains, significantly altering their substrate specificities. The identified subdomain of the HrmO-(β -Me)Phe₃ A domain was replaced with those from other NRPS A domains in the pathway based on putative recombination points. All hybrid enzymes were active and recognised their cognate amino acid substrates to generate the anticipated products, though similar swaps with subdomains from the calcium-dependent antibiotic (CDA) biosynthesis were found to be inactive (Crüsemann *et al.*, 2013). A similar logic was applied to the gramicidin biosynthetic pathway - out of nine chimeric constructs derived from the GrsA-A NRPS designed to alter the specificity of the phenylalanine-specific initiation module A domain, four activated their new anticipated amino acid substrate when assayed (Kries *et al.*, 2015). This method would avoid key residues identified as being important for MbtH-like domain interaction (Herbst *et al.*, 2013; Miller *et al.*, 2016). A sensible proof of concept for either of these approaches would be to alter the A₂ domain specificity to L-threonine, given that this was the amino acid substrate predicted by NRPSpredictor2 for this domain (Röttig *et al.*, 2011). An A domain involved in the biosynthesis of other β -lactone NPs isolated by Wells *et al.* (1982) would also appear to activate and tether L-threonine for cyclisation, if indeed these NPs are assembled by an analogous NRPS to **1**. If their producers' genome sequences can be obtained, it would allow the identification of the necessary A subdomain for swapping in.

If synthetic biology approaches can be applied successfully, it opens the door to efforts to design small molecule inhibitors of new enzyme targets. This potential is highlighted in the work of Sieber *et al.*, who found that a series of synthetic *trans*- β -lactone probes could specifically bind active site serine residues in caseinolytic protein proteases (ClpPs) (Böttcher and Sieber, 2008a). ClpP-binding led to attenuated extracellular virulence factor production in both *S. aureus* and *L. monocytogenes* (Böttcher and Sieber, 2008a; 2008b; 2009a; Gersch *et al.*, 2013). β -

Lactone-binding was found to be highly specific, and by altering the C-2 and C-3 side chains the inhibitors could be further structurally refined to produce more potent inhibitors (Böttcher and Sieber, 2009b; Weinandy *et al.*, 2014). Computational modelling would allow for the rational design of new β -lactone molecules with the physiochemical properties necessary to interact with novel biological targets (Rodrigues *et al.*, 2016); side chains could be selected that would best position the reactive β -lactone ring in close proximity to nucleophilic active site residues in the desired protein target. The use of a synthetic biological approach to generate new β -lactones would not incur the same costs associated with low-yielding reactions, and expensive chemical substrates and reaction conditions associated with chemical syntheses. Aside from the tractability of the system to manipulation, the only other drawback is the prerequisite for a β -hydroxy acid substrate for β -lactone ring formation with a suitable C $_{\alpha}$ group to introduce variable C-2 side chains.

9.3 Concluding remarks

This study investigating the biosynthesis and bioactivity of the antibiotic **1** has led to the identification of a new subfamily of PLP-dependent enzymes, the L-TTAs. These rare enzymes have great potential for the preparation of L-*threo*- β -OH- α -AAs for synthetic biology and chemistry approaches. Characterisation of individual genes within the **1** BGC has led to the identification of new queries for future genome mining efforts applying both mechanism- and redundancy-guided approaches. These should allow the specific targeting of uncharted NP chemical space to identify novel structures with potentially novel functions, and the identification of NPs with desired bioactivities, respectively. The modular nature of the **1** biosynthetic pathway lends itself to the rational design of synthetic β -lactone inhibitors of important protein targets. **1** was itself shown to be a potent inhibitor of ThrRSs in both Gram positive- and -negative bacteria, and likely represents a novel inhibitory mode of action for this enzyme target.

Bibliography

Bibliography

- Abbott, J. C., Aanensen, D. M. & Bentley, S. D.** WebACT: an online genome comparison suite. *Methods Mol. Bio.* **395**, 3422-3423 (2007).
- Achkar, J., Xian, M., Zhao, H. & Frost, J. W.** Biosynthesis of phloroglucinol. *J. Am. Chem. Soc.* **127**, 5332-5333 (2005).
- Ackerley, B. J., Rubin, E. J., Novick, V. L., Amaya, K., Judson, N. & Mekalanos, J.** A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. USA* **99**, 966-971 (2002).
- Ackerley, D. F. & Lamont, I. L.** Characterization and genetic manipulation of peptide synthetases in *Pseudomonas aeruginosa* PA01 in order to generate novel pyoverdines. *Chem. Biol.* **11**, 971-980 (2004).
- Aggarwal, R., Caffrey, P., Leadlay, P. F., Smith, C. J. & Staunton, J.** The thioesterase domain of the erythromycin-producing polyketide synthase: mechanistic studies *in vitro* to investigate its mode of action and substrate specificity. *J. Chem. Soc. Chem. Comm.* **15**, 1519-1520 (1995).
- Aguilera, S., López-López, K., Nieto, Y., Garcidueñas-Piña, R., Hernández-Guzmán, G., Hernández-Flores, J. L., Murillo, J. & Alvarez-Morales, A.** Functional characterization of the gene cluster from *P. syringae* pv. *phaseolicola* NPS3121 involved in synthesis of phaseolotoxin. *J. Bacteriol.* **189**, 2834-2843 (2007).
- Ahuja, E. G., Janning, P., Mentel, M., Graebisch, A., Breinbauer, R., Hiller, W., Costisella, B., Thomashow, L. S., Mavrodi, D. V. & Blankenfeldt, W.** PhzA/B catalyzes the formation of the tricycle in phenazine biosynthesis. *J. Am. Chem. Soc.* **130**, 17053-17061 (2008).
- Alanjary, M., Kronmiller, B., Adamek, M., Blin, K., Weber, T., Huson, D., Philmus, B. & Ziemert, N.** The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res.* **45**, W42-W48 (2017).
- Alexander, D. B & Zuberer, D. A.** Use of chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria. *Biol. Fertil. Soils* **12**, 39-45 (1991).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J.** Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403-410 (1990).
- Amann, R. I., Ludwig, W. & Schleifer, K. -H.** Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143-169 (1995).
- Anand, S., Prasad, M. V. R., Yadav, G., Kumar, N., Shehara, J., Ansari, M. Z. & Mohanty, D.** SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res.* **38**, W487-W496 (2010).

- Andexer, J. N., Kendrew, S. G., Nur-e-Alam, M., Lazos, O., Foster, T. A., Zimmerman, A. -S., Warneck, T. D., Suthar, D., Coates, N. J. & others.** Biosynthesis of the immunosuppressants FK506, FK520, and rapamycin involves a previously undescribed family of enzymes acting on chorismatase. *Proc. Natl. Acad. Sci. USA* **108**, 4776-4781 (2011).
- Andrews, F. H. & McLeish, M. J.** Using site-saturation mutagenesis to explore mechanism and substrate specificity in thiamine diphosphate-dependent enzymes. *FEBS J.* **280**, 6395-6411 (2013).
- Anthoni, U., Christophersen, C., Nielsen, P. H., Gram, L. & Petersen, B. O.** Pseudomonine, an isoxazolidone with siderophoric activity from *Pseudomonas fluorescens* AH2 isolated from Lake Victorian Nile perch. *J. Nat. Prod.* **58**, 1786-1789 (1995).
- Aoyagi, T., Tobe, H., Kojima, F., Hamada, M., Takeuchi, T. & Umezawa, H.** Amastatin, an inhibitor of aminopeptidase-A, produced by actinomycetes. *J. Antibiot.* **31**, 636-638 (1978).
- Arai, T. & Kino, K.** A novel L-amino acid ligase is encoded by a gene in the phaseolotoxin biosynthetic gene cluster from *Pseudomonas syringae* pv. *phaseolicola* 1448A. *Biosci. Biotechnol. Biochem.* **72**, 3048-3050 (2008).
- Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L. & others.** Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108-160 (2013).
- Aron, Z. D., Dorrestein, P. C., Blackhall, J. R., Kelleher, N. L. & Walsh, C. T.** Characterization of a new tailoring domain in polyketide biogenesis: The amine transferase domain of McyA in the mycosubtilin gene cluster. *J. Am. Chem. Soc.* **127**, 14986-14987 (2005).
- Arrebola, E., Carrión, V. J., Cazorla, F. M., Pérez-García, A., Murillo, J. & de Vicente, A.** Characterisation of the *mgo* operon in *Pseudomonas syringae* pv. *Syringae* UMAF0158 that is required for mangotoxin production. *BMC Microbiol.* **12**:10 (2012).
- Asai, A., Hasegawa, A., Ochiai, K., Yamashita, Y. & Mizukami, T.** Belactosin A, a novel antitumor antibiotic acting on cyclin/CDK mediated cell cycle regulation, produced by *Streptomyces* sp.. *J. Antibiot.* **53**, 81-83 (2000).
- Bagwell, C. L., Moloney, M. G. & Thompson, A. L.** On the antibiotic activity of oxazolomycin. *Bioorg. Med. Chem. Lett.* **18**, 4081-4086 (2008).
- Bai, T., Zhang, D., Lin, S., Long, Q., Wang, Y., Ou, H., Kang, Q., Deng, Z., Liu, W. & Tao, M.** Operon for the biosynthesis of lipstatin, the beta-lactone inhibitor of human pancreatic lipase. *Appl. Environ. Microbiol.* **80**, 7473-7483 (2014).
- Balado, M., Osorio, C. R. & Lemos, M. L.** Biosynthetic and regulatory elements involved in the production of the siderophore vanchrobactin in *Vibrio anguillarum*. *Microbiology* **154**, 1400-1413 (2008).
- Balibar, C. J., Vaillancourt, F. H. & Walsh, C. T.** Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem. Biol.* **12**, 1189-1200 (2005).

- Balskus, E. P. & Walsh, C. T.** Investigating the initial steps in the biosynthesis of cyanobacterial sunscreen scytonemin. *J. Am. Chem. Soc.* **130**, 15260-15261 (2008).
- Baltrus, D. A., Nishimura, M. T., Romanchuk, A., Chang, J. H., Mukhtar, M. S., Cherkis, K., Roach, J., Grant, S. R., Jones, C., D. & Dangl, J. L.** Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathog.* **7**, e1002132 (2011).
- Baltz, R.** Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol.* **8**, 557-563 (2008).
- Baltz, R.** Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **38**, 1747-1760 (2011).
- Bandara, H. M. D., Kennedy, D. P., Akin, E., Incarvito, C. D. & Burdette, S. C.** Photoinduced release of Zn²⁺ with ZinCleave-1: a nitrobenzyl-based caged complex. *Inorg. Chem.* **48**, 8445-8455 (2009).
- Bangera, M. G. & Thomashow, L. S.** Identification and characterisation of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from *Pseudomonas fluorescens* Q2-87. *J. Bacteriol.* **181**, 3155-3163 (1999).
- Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S. & Milo, R.** The moderately efficient enzyme: evolutionary and physiochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402-4410 (2011).
- Bar-Even, A. & Tawfik, D. S.** Engineering specialized metabolic pathways – is there a room for enzyme improvements? *Curr. Opin. Biotechnol.* **24**, 310-319 (2013).
- Barnard-Britson, S., Chi, X., Nonaka, K., Spork, A. P., Tibrewal, N., Goswami, A., Pahari, P., Ducho, C., Rohr, J. & van Lanen, S. G.** Amalgamation of nucleosides and amino acids in antibiotic biosynthesis: discovery of an L-threonine:uridine-5'-aldehyde transaldolase. *J. Am. Chem. Soc.* **134**, 18514-18517 (2012).
- Bauer, J. S., Ghequire, M. G. K., Nett, M., Josten, M., Sahl, H. -G., De Mot, R. and Gross, H.** Biosynthetic origin of the antibiotic pseudopyronines A and B in *Pseudomonas putida* BW11M1. *ChemBioChem* **16**, 2491-2497 (2015).
- Bender, C. L., Alarcón-Chaidez, F. & Gross, D. C.** *Pseudomonas syringae* phytotoxins: Mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiol. Mol. Biol. Rev.* **63**, 266-292 (1999).
- Bentley, S. D., Chater, K.F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., & others.** Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147 (2002).
- Bérdy, J.** Bioactive microbial metabolites. *J. Antibiot.* **58**, 1-26 (2005).
- Bérdy, J.** Thoughts and facts about antibiotics: where we are now and where we are heading. *J. Antibiot.* **65**, 385-395 (2012).
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E.** The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242 (2000).

- Bian, X., Plaza, A., Yan, F., Zhang, Y. & Müller, R.** Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminide derivatives *in vivo*. *Biotechnol. Bioeng.* **112**, 1343-1353 (2015).
- Bibb, M. J.** Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **8**, 208-215 (2005).
- Bickel, H., Mertens, P., Prelog, V., Seibl, J. & Wasler, A.** Constitution of ferrimycin A₁. *Antimicrob. Agents Chemother.* **5**, 951-957 (1965).
- Bisang, C., Long, P. F., Cortés, J., Westcott, J., Crosby, J., Matharu, A. L., Cox, R. J., Simpson, T. J., Staunton, J. & Leadlay, P. F.** A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **30**, 502-505 (1999).
- Blanc, V., Gill, P., Bamas-Jacques, N., Lorenzon, S., Zagorec, M., Schleuniger, J., Bisch, D., Blanche, F., Debussche, L., Cruzet, J. & Thibaut, D.** Identification and analysis of the genes from *Streptomyces pristinaespiralis* encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I. *Mol. Microbiol.* **23**, 191-202 (1997).
- Blankenfeldt, W., Kuzin, A. P., Skarina, T., Korniyenko, Y., Tong, L., Bayer, P., Janning, P., Thomashow, L. S. & Mavrodi, D. V.** Structure and function of the phenazine biosynthetic protein PhzF from *Pseudomonas fluorescens*. *Proc. Natl. Acad. Sci. USA* **101**, 16431-16436 (2004).
- Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kaustar, S. A., Suarez Duran, H. G., de Los Santos, E. L. C., Kim, H. U. & others.** antiSMASH 4.0 – improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36-W41 (2017).
- Blin, K., Medema, M. H., Kottmann, R., Lee, S. Y. and Weber, T.** The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* **45**, D555-D559 (2017).
- Blumer, C. & Haas, D.** Mechanism, regulation, and ecological role of bacterial cyanide biosynthesis. *Arch. Microbiol.* **173**, 170-177 (2000).
- Böckmann, M., Taraz, K. & Budzikiewicz, H.** Biogenesis of the pyoverdine chromophore. *Z. Naturforsch.* **52c**, 319-324 (1997).
- Boll, B., Taubitz, T. & Heide, L.** Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J. Biol. Chem.* **286**, 36281-36290 (2011).
- Borchardt, J. K.** The beginnings of drug therapy: Ancient mesopotamian medicine. *Drug News Perspec.* **15**, 187-192 (2002).
- Böttcher, T. & Sieber, S. A.** β -lactones as privileged structures for the active-site labeling of versatile bacterial enzyme classes. *Angew. Chem. Intl. Ed.* **47**, 4600-4603 (2008).
- Böttcher, T. & Sieber, S. A.** β -lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of *Staphylococcus aureus*. *J. Am. Chem. Soc.* **130**, 14400-14401 (2008).

- Böttcher, T. & Sieber, S. A.** β -lactones decrease the intracellular virulence of *Listeria monocytogenes* in macrophages. *ChemMedChem* **4**, 1260-1263 (2009).
- Böttcher, T. & Sieber, S. A.** Structurally refined β -lactones as potent inhibitors of devastating bacterial virulence factors. *ChemBioChem* **10**, 663-666 (2009).
- Böttcher, T. & Sieber, S. A.** β -Lactams and β -lactones as activity-based probes in chemical biology. *Med. Chem. Comm.* **3**, 408-417 (2012).
- Boyle, V. J., Fancher, M. E. & Ross Jr., R. W.** Rapid, modified Kirby-Bauer susceptibility test with single, high-concentration antimicrobial disks. *Antimicrob. Agents. Chemother.* **3**, 418-424 (1973).
- Bradford, M. M.** A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254 (1976).
- Brakhage, A. A.** Regulation of fungal secondary metabolism. *Nat. Rev. Microbiol.* **11**, 21-32 (2013).
- Braun, V., Günther, K., Hantke, K. & Zimmermann, L.** Intracellular activation of albomycin in *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* **156**, 308-315 (1983).
- Braun, V., Killmann, H. & Benz, R.** Energy-coupled transport through the outer membrane of *Escherichia coli* small deletions in the gating loop convert the FhuA transport protein into a diffusion channel. *FEBS Lett.* **346**, 59-64 (1994).
- Bredenbruch, F., Nitz, M., Wray, V., Morr, M., Müller, R. & Häussler, S.** Biosynthetic pathway of *Pseudomonas aeruginosa* 4-hydroxy-2-alkylquinolines. *J. Bacteriol.* **187**, 3630-3636 (2005).
- Breidenstein, E. B., de la Fuente-Nunez, C. & Hancock, R. E.** *Pseudomonas aeruginosa*: all roads lead to resistance. *Trends Microbiol.* **19**, 419-426 (2011).
- Brendel, N., Partida-Martinez, L. P., Scherlach, K. & Hertweck, C.** A cryptic PKS-NRPS gene locus in the plant commensal *Pseudomonas fluorescens* Pf-5 codes for the biosynthesis of an antimetabolic rhizoxin complex. *Org. Biomol. Chem.* **5**, 2211-2213 (2007).
- Brooks, D. M., Bender, C. L. & Kunkel, B. N.** The *Pseudomonas syringae* phytotoxin coronatine promotes virulence by overcoming salicylic acid-dependent defences in *Arabidopsis thaliana*. *Mol. Plant Pathol.* **6**, 629-639 (2005).
- Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., Dodson, R. J., Deboy, R. T., Durkin, A. S. & others.** The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. *Proc. Natl. Acad. Sci. USA* **100**, 10181-10186 (2003).
- Bugg, T. D. H., Wright, G. D., Dutka-Malen, S., Arthur, M., Courvalin, P. & Walsh, C. T.** Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: Biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry* **30**, 10408-10415 (1991).

- Bulkley, D., Innis, C. A., Blaha, G. & Steitz, T. A.** Revisiting the structures of several antibiotics bound to the bacterial ribosome. *Proc. Natl. Acad. Sci. USA* **107**, 17158-17163 (2010).
- Cai, W., Goswami, A., Yang, Z., Liu, X., Green, K. D., Barnard-Britson, S., Baba, S., Funabashi, M., Nonaka, K. & others.** The biosynthesis of capuramycin-type antibiotics: identification of the A-102395 biosynthetic gene cluster, mechanism of self-resistance, and formation of uridine-5'-carboxamide. *J. Biol. Chem.* **290**, 13710-13724 (2015).
- Calcott, M. J., Owen, J. G., Lamont, I. L. & Ackerley, D. F.** Biosynthesis of novel pyoverdines by domain substitution in a nonribosomal peptide synthetase of *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.* **80**, 5723-5731 (2014).
- Caldara-Festin, G., Jackson, D. R., Barajas, J. F., Valentic, T. R., Patel, A. B., Aguilar, S., Nguyen, M., Vo, M., Khanna, A. & others.** Structural and functional analysis of two di-domain aromatase/cyclases from type II polyketide synthases. *Proc. Natl. Acad. Sci. USA*. **112**, E6844-E6851 (2015).
- Calfee, M. W., Coleman, J. P. & Pesci, E. C.** Interference with *Pseudomonas* quinolone signal synthesis inhibits virulence factor expression by *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA* **98**, 11633-11637 (2001).
- Camilli, A. & Bassler, B. L.** Bacterial small-molecule signalling pathways. *Science* **311**, 1113-1116 (2006).
- Capra, J. A. & Singh, M.** Predicting functionally important residues from sequence conservation. *Bioinformatics* **23**, 1875-1882 (2007).
- Carrión, V. J., Arrebola, E., Cazorla, F. M., Murillo, J. & de Vicente, A.** The *mbo* operon is specific and essential for biosynthesis of mangotoxin in *Pseudomonas syringae*. *PLoS ONE* **7**, e36709 (2012).
- Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J.** ACT: the Artemis comparison tool. *Bioinformatics* **21**, 3422-3423 (2005).
- Carver, T. J., Thomson, N., Bleasby, A., Berriman, M. & Parkhill, J.** DNAPlotter: circular and linear interactive genome visualisation. *Bioinformatics* **25**, 119-120 (2009).
- Challis, G.** Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* **154**, 1555-1569 (2008).
- Chan, A. N., Shiver, A. L., Wever, W. J., Ravzi, S. Z. A., Traxler, M. F. & Li, B.** Role for dithiopyrrolones in disrupting bacterial metal homeostasis. *Proc. Natl. Acad. Sci. USA* **114**, 2717-2722 (2017).
- Chen, H., Guo, Z. & Liu, H. -W.** Biosynthesis of yersinirose: attachment of the two-carbon branched-chain is catalysed by a thiamine pyrophosphate-dependent flavoprotein. *J. Am. Chem. Soc.* **120**, 11796-11797 (1998).
- Chen, D. & Frey, P. A.** Identification of lysine 346 as a functionally important residue for pyridoxal-5'-phosphate binding and catalysis in lysine 2,3-aminomutase from *Bacillus subtilis*. *Biochemistry* **40**, 596-602 (2001).

- Chen, H. & Walsh, C. T.** Coumarin formation in novobiocin biosynthesis: β -hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI. *Chem. Biol.* **8**, 301-312 (2001).
- Chen, H., Hubbard, B. K., O'Connor, S. E. & Walsh, C. T.** Formation of β -hydroxy histidine in the biosynthesis of nikkomycin antibiotics. *Chem. Biol.* **9**, 103-112 (2002).
- Cheng, L., Chen, W., Zhai, L., Xu, D., Huang, T., Lin, S., Zhou, X. and Deng, Z.** Identification of the gene cluster involved in muraymycin biosynthesis from *Streptomyces* sp. NRRL 30471. *Mol. Biosyst.* **7**, 920-927 (2011).
- Chi, X., Baba, S., Tibrewal, N., Funabashi, M., Nonaka, K. & van Lanen, S. G.** The muraminomicin biosynthetic gene cluster and enzymatic formation of the 2-deoxyaminoribosyl appendage. *Med. Chem. Comm.* **4**, 239-243 (2013).
- Choi, K. -H. & Schweizer, H. P.** mini-Tn7 insertion in bacteria with single *attTn7* sites: example *Pseudomonas aeruginosa*. *Nat. Protoc.* **1**, 153-161 (2006)
- Choi, Y. S., Zhang, H., Brunzelle, J. S., Nair, S. K. & Zhao, H.** *In vitro* reconstitution and crystal structure of *p*-aminobenzoate *N*-oxygenase (AurF) involved in aureothin biosynthesis. *Proc. Natl. Acad. Sci. USA* **105**, 6858-6863 (2008).
- Chooi, Y. -H., Cacho, R. & Tang, Y.** Identification of the viridicatumatoxin and Griseofulvin gene clusters from *Penicillium aethiopicum*. *Chem. Biol.* **17**, 483-494 (2010).
- Christenson, J. K., Richman, J. E., Jensen, M. R., Neufeld, J. Y., Wilmot, C. M. & Wackett, L. P.** β -lactone synthetase found in the olefin biosynthesis pathway. *Biochemistry* **56**, 348-351 (2017).
- Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Brown, L. C. W., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M. & others.** Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412-421 (2014).
- Clarke, E. T., Braun, V., Winkelmann, G., Tari, L. W. & Vogel, H. J.** X-ray crystallographic structures of the *Escherichia coli* periplasmic protein FhuD bound to hydroxamate-type siderophores and the antibiotic albomycin. *J. Biol. Chem.* **277**, 13966-13972 (2002).
- Clarke, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W.** GenBank. *Nucleic Acids Res.* **44**, D67-D72 (2016).
- Cochrane, R. V. K., Norquay, A. K. & Vederas, J. C.** Natural products and their derivatives as tRNA synthetase inhibitors and antimicrobial agents. *Med. Chem. Comm.* **7**, 1535-1545 (2016).
- Coleman, J. P., Hudson, L. L., McKnight, S. L., Farrow, J. M., Calfee, M. W., Lindsey, C. A. & Pesci, E. C.** *Pseudomonas aeruginosa* PqsA is an anthranilate-coenzyme A ligase. *J. Bacteriol.* **190**, 1247-1255 (2008).
- Cooper, S. M., Cox, R. J., Crosby, J., Hothersall, J., Laosripaiboon, W., Simpson, T. J. & Thomas, C. M.** Mupirocin W, a novel pseudomonic acid produced by targeted mutation of the mupirocin biosynthetic gene cluster. *Chem. Commun.* **7**, 1179-1181 (2005).

- Contestabile, R., Paiardini, A., Pascarella, S., di Salvo, M. L., D'Aguanno, S. & Cossa, F.** L-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase. A subgroup of strictly related enzymes specialized for different functions. *Eur. J. Biochem* **268**, 6508-6525 (2001).
- Cornelis, P.** Iron uptake and metabolism in pseudomonads. *Appl. Microbiol. Biotechnol.* **86**, 1637-1645 (2010).
- Costacurta, A. & Vanderleyden, J.** Synthesis of phytohormones by plant-associated bacteria. *Crit. Rev. Microbiol.* **21**, 1-18 (2008).
- Costa, R., van Aarle, I. M., Mendes, R. & van Elsas, J. D.** Genomics for pyrrolnitrin biosynthetic loci: evidence for conservation and whole-operon mobility within Gram-negative bacteria. *Environ. Microbiol.* **11**, 159-175 (2009).
- Costelloe, S. J., Ward, J. M. & Dalby, P. A.** Evolutionary analysis of the TPP-dependent enzyme family. *J. Mol. Evol.* **66**, 36-49 (2008).
- Cox, C. D., Rinehart, K. L., Moore, M. L. & Cook, J. C.** Pyochelin: Novel structure of an iron-chelating growth promoter for *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA* **78**, 4256-4260 (1981).
- Cox, J. & Mann, M.** MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).
- Crone, W. J. K., Vior, N. M., Santos-Aberturas, J., Schmitz, L. G., Leeper, F. J. & Truman, A. W.** Dissecting bottromycin biosynthesis using comparative untargeted metabolomics. *Angew. Chem. Int. Ed.* **55**, 9639-9643 (2016).
- Crosa, J. H. & Walsh, C. T.** Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **66**, 223-249 (2002).
- Crüseman, M., Kohlhaas, C. & Piel, J.** Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chem. Sci.* **4**, 1041-1045 (2013).
- Cruz-Morales, P., Kopp, J. F., Martínez-Guerrero, C., Yáñez-Guerra, L. A., Selem-Mojica, N., Ramos-Aboites, H., Feldmann, J. & Barona-Gómez, F.** Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model Streptomyces. *Genome Biol. Evol.* **8**, 1906-1916 (2016).
- Cubillos-Ruiz, A., Berta-Thompson, J. W., Becker, J. W., van der Donk, W. A. and Chisolm, S. W.** Evolutionary radiation of lanthipeptides in marine cyanobacteria. *Proc. Natl. Acad. Sci. USA* **114**, E5424-E5433 (2017).
- Cundliffe, E. & Demain, A. L.** Avoidance of suicide in antibiotic-producing microbes. *J. Ind. Microbiol. Biotechnol.* **37**, 643-672 (2010).
- Davies, J. & Davies, D.** Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**, 417-433 (2010).
- Dawlaty, J., Zhang, X., Fischbach, M. A. & Clardy, J.** Dapdiamides, tripeptide antibiotics formed by unconventional amide ligases. *J. Nat. Prod.* **73**, 441-446 (2010).

- D'Costa, V. M., King, C. E., Kalan, M., Mariya, M., Sung, W. W. L., Schwartz, C., Froese, D., Zazula, G., Calmels, F. & others.** Antibiotic resistance is ancient. *Nature* **477**, 457-461 (2011).
- De Bruijn, I., de Kock, M. J. D., de Waard, P., van Beek, T. A. & Raaijmakers, J. M.** Massetolide A biosynthesis in *Pseudomonas fluorescens*. *J. Bacteriol.* **190**, 2777-2789 (2008).
- De Lorenzo, V., Bindereif, A., Paw, B. H. & Neilands, J. B.** Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in *Escherichia coli* K-12. *J. Bacteriol.* **165**, 570-578 (1986).
- De Pascale, G., Nazi, I., Harrison, P. H. & Wright, G. D.** Beta-Lactone natural products and derivatives inactivate homoserine transacetylase, a target for antimicrobial agents. *J. Antibiot.* **64**, 483-487 (2011).
- Dejong, C. A., Chen, G. M., Li, H., Johnston, C. W., Edwards, M. R., Rees, P. N., Skinnider, M. A., Webster, A. L. H. & Magarvey, N. A.** Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat. Chem. Biol.* **12**, 1007-1014 (2016).
- Demain, A. L.** Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol., Biotechnol.* **14**, 185-201 (2014).
- Denard, C. A., Ren, H. & Zhao, H.** Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Chem. Biol.* **25**, 55-64 (2015).
- Deng, H., Cross, S. M., McGlinchey, R. P., Hamilton, J. T. G. & O'Hagan, D.** *In vitro* reconstituted biotransformation of 4-fluorothreonine from fluoride ion: application of the fluorinase. *Chem. Biol.* **15**, 1268-1276 (2008).
- Devine, J. H., Shadel, G. S. & Baldwin, T. O.** Identification of the operator of the *lux* regulon from the *Vibrio fischeri* strain ATCC7744. *Proc. Natl. Acad. Sci. USA* **86**, 5688-5692 (1989).
- Dick, L. R., Cruikshank, A. A., Grenier, L., Melandri, F. D., Nunes, S. L. & Stein, R. L.** Mechanistic studies on the inactivation of the proteasome by lactacystin: a central role for *clasto*-lactacystin β -lactone. *J. Biol. Chem.* **271**, 7273-7276 (1996).
- Dietrich, L. E. P., Teal, T. K., Price-Whelan, A. & Newman, D. K.** Redox-active antibiotics control gene expression and community behaviour in divergent bacteria. *Science* **321**, 1203-1206 (2008).
- Dietrich, L. E. P., Okegbe, C., Price-Whelan, A., Saktah, H., Hunter, R. C. and Newman, D. K.** Bacterial community morphogenesis is intimately linked to the intracellular redox state. *J. Bacteriol.* **195**, 1371-1380 (2013).
- Di Salvo, M. L., Budisa, N. & Contestabile, R.** PLP-dependent enzymes: A powerful tool for metabolic synthesis of non-canonical amino acids. *Molekulare Entwicklung und Kontrolle*, Beilstein Symposium (2013).
- DiSpirito, A. A., Zahn, J. A., Graham, D. W., Kim, H. J., Larive, C. K., Derrick, T. S., Cox, C. D. & Taylor, A. B.** Copper-binding compounds from *Methylosinus trichosporium* OB3b. *J. Bacteriol.* **180**, 3606-3613 (1998).

- Donia, M. S., Hathaway, B. J., Sudek, S., Haygood, M. G., Rosovitz, M. J., Ravel, J. & Schmidt, E. W.** Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat. Chem. Biol.* **2**, 729-735 (2006).
- D'Onofrio, A., Crawford, J. M., Stewart, E. J., Witt, K., Gavrish, E., Epstein, S., Clardy, J. & Lewis, K.** Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* **17**, 254-264 (2010).
- Doong, R. -A. & Lei, W. -G.** Solubilization and mineralization of polycyclic aromatic hydrocarbons by *Pseudomonas putida* in the presence of surfactant. *J. Hazard. Mater.* **96**, 15-27 (2003).
- Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K. -S., Haines, R. R., Tchalukov, K. A., Labeda, D. P., Kelleher, N. L. & Metcalf, W. M.** A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **10**, 963-968 (2014).
- Drake, E. J., Nicolai, D. A. & Gulick, A. M.** Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chem. Biol.* **13**, 409-419 (2006).
- Drake, E. J., Miller, B. R., Shi, C., Tarrasch, J. T., Sundlov, J. A., Allen, C. L., Skinotis, G., Aldrich, C. C. & Gulick, A. M.** Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235-238 (2016).
- Drees, S. L. & Fetzner, S.** PsqE of *Pseudomonas aeruginosa* acts as pathway-specific thioesterase in the biosynthesis of alkylquinolone signaling molecules. *Chem. Biol.* **22**, 611-618 (2015).
- Drees, S. L., Chan, L., Prasetya, F., Saleem, M., Dreveny, I., Williams, P., Hennecke, U., Emsley, J. & Fetzner, S.** PqsBC, a condensing enzyme in the biosynthesis of the *Pseudomonas aeruginosa* quinolone signal. *J. Biol. Chem.* **291**, 6610-6624 (2016).
- Du, L., Sánchez, C. & Shen, B.** Hybrid peptide-polyketide natural products: biosynthesis prospects toward engineering novel molecules. *Metabol. Engineer.* **3**, 78-95 (2001).
- Du, L. & Lou, L.** PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **27**, 255-278 (2010).
- Du, Y. -L., Alkhalaf, L. M. & Ryan, K. S.** *In vitro* reconstitution of indolmycin biosynthesis reveals the molecular basis of oxazolinone assembly. *Proc. Natl. Acad. Sci. USA* **112**, 2717-2722 (2015).
- Duberne, J. -F., Coppoolse, E. R., Stiekema, W. J. & Bloemberg, G. V.** Genetic and functional characterization of the gene cluster directing the biosynthesis of putisolvin I and II in *Pseudomonas putida* strain PCL1445. *Microbiology* **154**, 2070-2083 (2008).
- Dücker, N., Baer, K., Simon, S., Gröger, H. & Hummel, W.** Threonine aldolases – screening, properties and applications in the synthesis of non-proteinogenic β -hydroxy- α -amino acids. *App. Microbiol. Biotechnol.* **88**, 409-424 (2010).
- Duggleby, R. G.** Domain relationships in thiamine diphosphate-dependent enzymes. *Acc. Chem. Res.* **39**, 550-557 (2006).

- Dunathan, H. C.** Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc. Natl. Acad. Sci. USA* **55**, 712-716 (1966).
- Duquesne, S., Destoumieux-Garzón, Peduzzi, J. & Rebuffat, S.** Microcins, gene-encoded antibacterial peptides from enterobacteria. *Nat. Prod. Rep.* **24**, 708-734 (2007).
- Edgar, R. C.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).
- Eisenbeis, S. & Hocker, B.** Evolutionary mechanism as a template for protein engineering. *J. Pept. Sci.* **16**, 538-544 (2010).
- Eisenreich, W., Kupfer, E., Weber, W. & Bacher, A.** Tracer studies with crude U-¹³C-lipid mixtures. Biosynthesis of the lipase inhibitor lipstatin. *J. Biol. Chem.* **272**, 867-874 (1997).
- Eisenreich, W., Kupfer, E., Stohler, P., Weber, W. & Bacher, A.** Biosynthetic origin of a branched chain analogue of the lipase inhibitor, lipstatin. *J. Med. Chem.* **46**, 4209-4212 (2003).
- Eliot, A. C. & Kirsch, J. F.** Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* **73**, 383-415 (2004).
- El-Sayed, A. K., Hothersall, J. & Thomas, C. M.** Quorum-sensing-dependent regulation of biosynthesis of the polyketide antibiotic mupirocin in *Pseudomonas fluorescens* NCIMB 10586. *Microbiology* **147**, 2127-2139 (2001).
- El-Sayed, A. K., Hothersall, J., Cooper, S. M., Stephens, E., Simpson, T. J. and Thomas, C. M.** Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* **10**, 419-430 (2003).
- Endicott, N. P., Lee, E. & Wencewicz, T. A.** Structural basis for xenosiderophore utilization by the human pathogen *Staphylococcus aureus*. *ACS Infect. Dis.* **3**, 542-553 (2017).
- Eppelmann, K., Stachelhaus, T. & Marahiel, M. A.** Exploitation of the selectivity-conferring code of the nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**, 9718-9726 (2002).
- Ernst, J. F., Bennett, R. L. & Rothfield, L. I.** Constitutive expression of the iron-enterochelin and ferrichrome uptake systems in a mutant strain of *Salmonella typhimurium*. **135**, 928-934 (1978).
- Eustáquio, A. S., Pojer, F., Noel, J. P. & Moore, B. S.** Discovery and characterisation of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **4**, 69-74 (2008).
- Eustáquio, A. S., McGlinchey, R. P., Liu, Y., Hazzard, C., Beer, L. L., Florova, G., Alhamadsheh, M. M., Kale, A. J., Kobayashi, Y., Reynolds, K. A. & Moore, B. S.** Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. *Proc. Natl. Acad. Sci. USA* **106**, 12295-12300 (2009).

- Fang, P., Yu, X., Jeong, J. S., Mirando, A., Chen, K., Chen, X., Kim, S., Francklyn, C. S. & Guo, M.** Structural basis for full-spectrum inhibition of translational functions on a tRNA synthetase. *Nat. Commun.* **6**, 6402 (2015).
- Farrow, J. M. & Pesci, E. C.** Two distinct pathways supply anthranilate as a precursor of the *Pseudomonas* quinolone signal. *J. Bacteriol.* **189**, 3425-3433 (2007).
- Fawaz, M. V., Topper, M. E. & Firestine, S. M.** The ATP-grasp enzymes. *Bioorg. Chem.* **39**, 185-191 (2011).
- Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R. & Fenical, W.** Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chem. Int. Ed.* **42**, 355-357 (2003).
- Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D. W. & Thomas, M. G.** MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry* **49**, 8815-8817 (2010).
- Fesko, K., Reisinger, C., Steinreiber, J., Weber, H., Schürmann, M. & Griengl, H.** Four types of threonine aldolases: similarities and differences in kinetics/thermodynamics. *J. Mol. Catal. B-Enzym.* **52-53**, 19-26 (2008).
- Fesko, K., Uhl, M., Steinreiber, J., Gruber, K. & Greingl, H.** Biocatalytic access to α,α -dialkyl- α -amino acids by a mechanism-based approach. *Angew. Chem. Int. Ed.* **49**, 121-124 (2010).
- Fesko, K. & Gruber-Khadjawi, M.** Biocatalytic methods for C-C bond formation. *Chem. Cat. Chem.* **5**, 1248-1272 (2013).
- Fesko, K.** Threonine aldolases: perspectives in engineering and screening the enzymes with enhanced substrate and stereo specificities. *Appl. Microbiol. Biotechnol.* **100**, 2579-2590 (2016).
- Finking, R. & Marahiel, M. A.** Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58**, 453-488 (2004).
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M. & others.** The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279-D285 (2016).
- Firn, R. D. & Jones, C. G.** The evolution of secondary metabolism – a unifying model. *Mol. Microbiol.* **37**, 989-994 (2000).
- Fisch, K. M., Gurgui, C., Heycke, N., van der Sar, S. A., Anderson, S. A., Webb, V. L., Taudien, S., Platzer, M., Rubio, B. K. & others.** Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.* **5**, 494-501 (2009).
- Fischbach, M. A. & Walsh, C. T.** Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468-3496 (2006).
- Fischbach, M. A., Walsh, C. T. & Clardy, J.** The evolution of gene collectives: How natural selection drives chemical innovation. *Proc. Natl. Acad. Sci. USA* **105**, 4601-4608 (2008).

- Fischbach, M. A. & Walsh, C. T.** Antibiotics for emerging pathogens. *Science* **325**, 1089-1093 (2009).
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A. & others.** Whole-genome sequencing random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512 (1995).
- Fleming, A.** On the antibacterial action of cultures of *Penicillium*, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.* **10**, 226-236 (1929).
- Frueh, D. P., Arthanari, H., Koglin, A., Vosburg, D. A., Bennett, A. E., Walsh, C. T. & Wagner, G.** Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **454**, 903-907 (2008).
- Fu, C., Auerbach, D., Li, Y., Scheid, U., Luxenburger, E., Garcia, R., Irschik, H. & Müller, R.** Solving the puzzle of one-carbon loss in ripostatin biosynthesis. *Angew. Chem. Int. Ed.* **56**, 2192-2197 (2017).
- Funabashi, M., Baba, S., Takatsu, T., Kizuka, M., Ohata, Y., Tanaka, M., Nonaka, K., Spork, A. P., Ducho, C. & others.** Structure-based gene targeting discovery of sphaerimycin, a bacterial translocase I inhibitor. *Angew. Chem. Int. Ed.* **52**, 11607-11611 (2013).
- Gallagher, L. A., McKnight, S. L., Kuznetsova, M. S., Pesci, E. C. & Manoil, C.** Functions required for extracellular quinolone signaling by *Pseudomonas aeruginosa*. *J. Bacteriol.* **184**, 6472-6480 (2002).
- Gao, S. -S., Hothersall, J., Wu, J., Murphy, A. C., Song, Z., Stephens, E. R., Thomas, C. M., Crump, M. P., Cox, R. J. & others.** Biosynthesis of mupirocin by *Pseudomonas fluorescens* NCIMB 10586 involves parallel pathways, *J. Am. Chem. Soc.* **136**, 5501-5507 (2014).
- Garneau-Tsodikova, S., Dorrestein, P. C., Kelleher, N. L. & Walsh, C. T.** Protein assembly line components in prodigiosin biosynthesis: characterisation of PigA,G,H,I,J. *J. Am. Chem. Soc.* **128**, 12600-12601 (2006).
- Garrido-Sanz, D., Arrebola, E., Martínez-Granero, F., García-Méndez, S., Muriel, C., Blanco-Romero, E., Martín, M., Rivilla, R. & Redondo-Nieto, M.** Classification of isolates from the *Pseudomonas fluorescens* complex into phylogenomic groups based in group-specific markers. *Front. Microbiol.* **8**, (2017).
- Garsin, D. A., Urbach, J., Huguet-Tapia, J. C., Peters, J. E. & Ausubel, F. M.** Construction of an *Enterococcus faecalis* Tn917-mediated-gene-disruption library offers insight into Tn917 insertion patterns. *J. Bacteriol.* **186**, 7280-7289 (2004).
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A.** Protein identification and analysis tools on the ExPASy server. (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005).
- Gay, P., Lecoq, D., Steinmetz, M., Berkelman, T. & Kado, C. I.** Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol.* **164**, 918-921 (1985).

- Gehring, A. M., Bradley, K. A. & Walsh, C. T.** Enterobactin biosynthesis in *Escherichia coli*: Isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* **36**, 8495-8503 (1997).
- Geoffroy, M. -C., Floquet, S., Métais, A., Nassif, X. & Pelicic, V.** Large-scale analysis of the meningococcus genome by gene disruption: resistance to complement-mediated lysis. *Genome Res.* **13**, 391-398 (2003).
- Gerard, J., Lloyd, R., Barsby, T., Haden, P., Kelly, M. T. & Andersen, R. J.** Massetolides A-H, antimycobacterial cyclic depsipeptides produced by two pseudomonads isolated from marine habitats. *J. Nat. Prod.* **60**, 223-229 (1997).
- Gerber, R., Lou, L. and Du, L.** A PLP-dependent polyketide chain releasing mechanism in the biosynthesis of mycotoxin fumonisins in *Fusarium verticillioides*. *J. Am. Chem. Soc.* **131**, 3148-3149 (2009).
- Gersch, M., Gut, F., Korotkov, V. S., Lehmann, J., Böttcher, T., Rusch, M., Hedberg, C., Waldman, H., Klebe, G. & Sieber, S. A.** The mechanism of caseinolytic protease (ClpP) inhibition. *Angew. Chem. Int. Ed.* **52**, 3009-3014 (2013).
- Gibson, D. G., Young, L., Chuang, R. -Y., Venter, J. G., Hutchinson III, C. A. & Smith, H. O.** Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-345 (2009).
- Glickmann, E., Gardan, L., Jacquet, S., Hussain, S., Elasri, M., Petit, A. & Dessaux, Y.** Auxin production is a common feature of most pathovars of *Pseudomonas syringae*. *Mol. Plant Microbe Interact.* **11**, 156-162 (1998).
- Goese, M., Eisenreich, W., Kupfer, E., Stohler, P., Weber, W. & Bacher, A.** Biosynthetic origin of hydrogen atoms in the lipase inhibitor lipstatin. *J. Biol. Chem.* **275**, 21192-21196 (2000).
- Goese, M., Eisenreich, W., Kupfer, E., Stohler, P., Weber, W., Leuenberger, H. G., & Bacher, A.** Biosynthesis of lipstatin. Incorporation of multiply deuterium-labeled (5Z,8Z)-tetradeca-5,8-dienoic acid and octanoic acid. *J. Org. Chem.* **66**, 4673-4678 (2001).
- Goldstein, D. S.** L-Dihydroxyphenylserine (L-DOPS): a norepinephrine prodrug. *Cardiovasc. Drug Rev.* **24**, 189-203 (2006).
- Gomez-Escribano, J. P. & Bibb, M. J.** Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207-215 (2011).
- Gomez-Escribano, J. P., Alt, S. & Bibb, M. J.** Next generation sequencing of actinobacteria for the discovery of novel natural products. *Mar. Drugs* **14**, E78 (2016).
- Goswami, A. & van Lanen, S. G.** Enzymatic strategies and biocatalysts for amide bond formation: tricks of the trade outside of the ribosome. *Mol. Biosyst.* **20**, 338-353 (2015).

- Greenspan, M. D., Yudkovitz, J. B., Lo, C. -Y. L., Chen, J. S., Alberts, A. W., Hunt, V. M., Chang, M. N., Yang, S. S., Thompson, K. L. & others.** Inhibition of hydroxymethylglutaryl-coenzyme A synthase by L-659,699. *Proc. Natl. Acad. Sci. USA* **84**, 7488-7492 (1987).
- Groll, M., Bochtler, M., Brandstetter, H., Clausen, T. & Huber, R.** Molecular machines for protein degradation. *ChemBioChem* **6**, 222-256 (2005).
- Groll, M., Huber, R. & Potts, B. C. M.** Crystal structures of salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of β -lactone ring opening and a mechanism for irreversible binding. *J. Am. Chem. Soc.* **128**, 5136-5141 (2006).
- Groll, M., Larionov, O. V., Huber, R. & de Meijere, A.** Inhibitor-binding mode of homobelactosin C to proteasomes: New insights into class I MHC ligand generation. *Proc. Natl. Acad. Sci. USA* **103**, 4576-4579 (2006).
- Gross, H. & Loper, J. E.** Genomics of secondary metabolite production by *Pseudomonas* spp.. *Nat. Prod. Rep.* **26**, 1408-1446 (2009).
- Grue-Sørensen, G. & Spenser, I. D.** Biosynthesis of ephedrine. *J. Am. Chem. Soc.* **110**, 3714-3715 (1988).
- Grue-Sørensen, G. & Spenser, I. D.** The biosynthesis of ephedrine. *Can. J. Chem.* **67**, 998-1009 (1989).
- Gu, W. & Schmidt, E. W.** Three principles of diversity-generating biosynthesis. *Acc. Chem. Res.* Article in press.
- Gulder, T. A. M. & Moore, B. S.** Salinosporamide natural products: Potent 20S proteasome inhibitors as promising cancer chemotherapeutics. *Angew. Chem. Int. Ed. Engl.* **49**, 9346-9367 (2010).
- Guttenberger, N., Blankenfeldt, W. & Breinbauer, R.** Recent developments in the isolation, biological function, biosynthesis, and synthesis of phenazine natural products. *Bioorg. Med. Chem.* Article in press.
- Haas, D. & Défago, G.** Biological control of soil-borne pathogens by fluorescent *Pseudomonads*. *Nat. Rev. Microbiol.* **3**, 307-319 (2005).
- Hadjithomas, M., Chen, I. -M. A., Chu, K., Ratner, A., Palaniappan, K., Szeto, E., Huang, J., Reddy, T. B. K., Cimermancic, P. & others.** IMG-ABC: A knowledge base to fuel discovery of biosynthetic gene clusters and novel secondary metabolites. *mBio* **6**, e00932 (2015).
- Hadváry, P., Sidler, W., Meister, W., Vetter, W. & Wolfer, H.** The lipase inhibitor tetrahydrolipstatin binds covalently to the putative active site serine of pancreatic lipase. *J. Biol. Chem.* **266**, 2021-2027 (1991).
- Hall, T. A.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT *Nucleic Acids Symposium Series* **41**, 95-98 (1999).
- Hammer, P. E., Hill, D. S., Lam, S. T., van Pée, K. -H. & Ligon, J. M.** Four genes from *Pseudomonas fluorescens* that encode the biosynthesis of pyrrolnitrin. *Appl. Environ. Microbiol.* **63**, 2147-2154 (1997).

- Hammer, P. E., Burd, W., Hill, D. S., Ligon, J. M. & van Pée, K. -H.** Conservation of the pyrrolnitrin biosynthetic gene cluster among six pyrrolnitrin-producing strains *FEMS Microbiol. Lett.* **180**, 39-44 (1999).
- Hammer, N. D. & Skaar, E. P.** Molecular mechanisms of *Staphylococcus aureus* iron acquisition. *Annu. Rev. Microbiol.* **65**, 129-147 (2011).
- Hanahan, D.** Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557-580 (1983).
- Hannauer, M., Barda, Y., Mislin, G. L. A., Shanzer, A., Schalk, I. J.** The ferrichrome uptake pathway involves an iron release mechanism with acylation of the siderophore and recycling of the modified desferrichrome. *J. Bacteriol.* **192**, 1212-1220 (2010).
- Harborne, J. B.** Dictionary of natural products. <http://dnp.chemnetbase.com> (Taylor & Francis, 2015).
- Harrison, S. J., Mainwaring, P., Price, T., Millward, M. J., Padrik, P., Underhill, C. R., Cannell, P. K., Reich, S. D., Trikha, M. & Spencer, A.** Phase I clinical trial of Marizomib (NPI-0052) in patients with advanced malignancies including multiple myeloma: Study NPI-0052-102 final results. *Clin. Cancer Res.* **22**, 4559-4566 (2016).
- Hartmann, A., Fiedler, H. -P. & Braun, V.** Uptake and conversion of the antibiotic albomycin by *Escherichia coli* K-12. *Eur. J. Biochem.* **99**, 517-524 (1979).
- Hassan, A. Q., Kirby, C. A., Zhou, W., Schuhmann, T., Kityk, R., Kipp, D. R., Baird, J., Chen, J., Chen, Y. & others.** The novolactone natural product disrupts the allosteric regulation of Hsp70. *Chem. Biol.* **22**, 87-97 (2015).
- He, J., Magarvey, N., Pirae, M. & Vining, L. C.** The gene cluster for chloramphenicol biosynthesis in *Streptomyces venezuelae* ISP5230 includes novel shikimate pathway homologues and a monomodular non-ribosomal peptide synthetase gene. *Microbiol.* **147**, 2817-2828 (2001).
- He, J. & Hertweck, C.** Biosynthetic origin of the rare nitroaryl moiety of the polyketide antibiotic aureothin: involvement of an unprecedented *N*-oxygenase. *J. Am. Chem. Soc.* **126**, 3694-3695 (2004).
- Heemstra, J. R., Walsh, C. T. & Sattely, E. S.** Enzymatic tailoring of ornithine in the biosynthesis of the *Rhizobium* cyclic trihydroxamate siderophore vicibactin. *J. Am. Chem. Soc.* **131**, 15317-15329 (2009).
- Helfrich, E. J. N. & Piel, J.** Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231-316 (2016).
- Henckel, T. & Zeeck, A.** Inthomycins, new oxazole-trienes from *Streptomyces* sp. *Liebigs Ann. Chem.* **1991**, 1541-1546 (1991).
- Herbert, R. B. & Knaggs, A. K.** The biosynthesis of the antibiotic obafluorin from *p*-aminophenylalanine in *Pseudomonas fluorescens*. *Tetrahedron Lett.* **29**, 6353-6356 (1988).
- Herbert, R. B. & Knaggs, A. K.** The biosynthesis of the *Pseudomonas fluorescens* antibiotic obafluorin from *p*-aminophenylalanine and glycine (glyoxylate). *Tetrahedron. Lett.* **31**, 7517-7520 (1990).

- Herbert, R. B. & Knaggs, A. K.** Biosynthesis of the antibiotic obafluorin from D-[U-¹³C] glucose and *p*-aminophenylalanine in *Pseudomonas fluorescens*. *J. Chem. Soc. Perkin Trans. 1* 103-107 (1992).
- Herbert, R. B. & Knaggs, A. K.** Biosynthesis of the antibiotic obafluorin from *p*-aminophenylalanine and glycine (glyoxylate). *J. Chem. Soc. Perkin Trans. 1* 109-113 (1992).
- Herbert, R. B., Wilkinson, B. & Ellames, G. J.** Preparation of (2*R*,3*S*)- β -hydroxy- α -amino acids by use of a novel *Streptomyces* aldolase as a resolving agent for racemic material. *Can. J. Chem.* **72**, 114-117 (1994).
- Herbst, D. A., Boll, B., Zocher, G., Stehle, T. & Heide, L.** Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. *J. Biol. Chem.* **288**, 1991-2003 (2013).
- Hernandez, M. E. & Newman, D. K.** Extracellular electron transfer. *Cell. Mol. Life Sci.* **58**, 1562-1571 (2001).
- Hertweck, C.** The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.* **48**, 4688-4716 (2009).
- Heuer, H., Krsek, M., Baker, P., Smalla, K. & Wellington E. M. H.** Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl. Environ. Microbiol.* **63**, 3233-3241 (1997).
- Hider, R. C. & Kong, X.** Chemistry and biology of siderophores. *Nat. Prod. Rep.* **27**, 637-657 (2009).
- Hoffmann, T., Krug, D., Hüttel, S. and Müller, R.** Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **86**, 10780-10788 (2014).
- Hopwood, D. A.** *Streptomyces* in nature and medicine: the antibiotic makers. Oxford University Press, New York (USA) (2007).
- Hothersall, J., Wu, J., Rahman, A. S., Shields, J. A., Haddock, J., Johnson, N., Cooper, S. M., Stephens, E. R., Cox, R. J. & others.** Mutational analysis reveals that all tailoring region genes are required for production of polyketide antibiotic mupirocin by *Pseudomonas fluorescens* – Pseudomonic acid B biosynthesis precedes pseudomonic acid A. *J. Biol. Chem.* **282**, 15451-15461 (2007).
- Huoang, T. T., Karkhoff-Schweizer, R. R., Kutchma, A. J. & Schweizer, H. P.** A broad-host-range Flp-*FRT* recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. *Gene* **28**, 77-86 (1998).
- Hur, G. H., Vickery, C. R. & Burkart, M. D.** Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **29**, 1074-1098 (2012).
- Hutter, B., Fischer, C., Jacobi, A., Schaab, C. & Loferer, H.** Panel of *Bacillus subtilis* reporter strains indicative of various modes of action. *Antimicrob. Agents Chemother.* **48**, 2588-2594 (2004).

- Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W. & Hauser, L. J.** Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**,119 (2010).
- Ikeda, Y., Kondo, S., Naganawa, H., Hattori, S., Hamada, M. & Takeuchi, T.** New triene- β -lactone antibiotics, triedimycins A and B. *J. Antibiot.* **44**, 453-455 (1991).
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., & Omura, S.** Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21**, 526-531 (2003).
- Ito, T. & Masubuchi, M.** Dereplication of microbial extracts and related analytical technologies. *J. Antibiot.* **67**, 353-360 (2014).
- Jensen, S. E.** Biosynthesis of clavam metabolites. *J. Ind. Microbiol. Biotechnol.* **39**, 1407-1419 (2012).
- Jiang, W., Hou, Y. & Inouye, M.** CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J. Biol. Chem.* **3**, 196-202 (1997).
- Jiang, L., Wang, L., Zhang, J., Liu, H., Hong, B., Tan, H. & Niu, G.** Identification of novel mureidomycin analogues via rational activation of a cryptic gene cluster in *Streptomyces roseosporus* NRRL 15998. *Sci. Rep.* **5**, 14111 (2015).
- Jiménez, J. I., Miñambres, B., García, J. L. & Díaz, E.** Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 824-841 (2002).
- Jin, B., Newton, S. M. C., Shao, Y., Jiang, X., Charbit, A. & Klebba, P. E.** Iron acquisition systems for ferric hydroxamates, haemin and haemoglobin in *Listeria monocytogenes*. *Mol. Microbiol.* **59**, 1185-1198 (2006).
- Johnston, C. W., Skinnider, M. A., Dejong, C. A., Rees, P. A., Chen, G. M., Walker, C. G., French, S., Brown, E. D., Bérdy, J., Liu, D. Y. & Magarvey, N. A.** Assembly and clustering of natural antibiotics guides target identification. *Nat. Chem. Biol.* **12**, 233-239 (2016).
- Jordan, F.** Current mechanistic understanding of thiamine diphosphate-dependent enzymatic reactions. *Nat. Prod. Rep.* **20**, 184-201 (2003).
- Ju, K. -S. & Parales, R. E.** Nitroaromatic compounds, from synthesis to biodegradation. *Microbiol. Mol. Biol. Rev.* **74**, 250-272 (2010).
- Jurkevitch, E., Hadar, Y., Chen, Y., Yakirevitch, P., Libman, J. & Shanzer, A.** Iron uptake and molecular recognition in *Pseudomonas putida*: a receptor mapping ferrioxamine B, coprogen B and their biomimetic analogues. *Microbiology* **140**, 1697-1703 (1994).
- Kadi, N. & Challis, G. L.** Siderophore biosynthesis: a substrate specificity assay for nonribosomal peptide synthetase-independent siderophore synthetases involving trapping of acyl-adenylate intermediates with hydroxylamine. *Methods Enzymol.* **458**, 431-457 (2009).

- Kahlmeter, G., Brown, D. F., Goldstein, F. W., MacGowan, A. P., Mouton, J. W., Odenholt, I., Rodloff, A., Soussy, C. J., Steinbakk, M., Soriano, F. & Stetsiouk, O.** European Committee on Antimicrobial Susceptibility Testing (EUCAST) technical notes of antimicrobial susceptibility testing. *Clin. Microbiol. Infect.* **12**, 501-503 (2006).
- Kang, J. G. & Park, C. Y.** Anti-obesity drugs: a review about their effects and safety. *Diabetes Metab. J.* **36**, 13-25 (2012).
- Kang, H. -S. & Kim, J. -P.** Ostalactones A-C, β - and ϵ -lactones with lipase inhibitory activity from the cultured basidiomycete *Stereum ostrea*. *J. Nat. Prod.* **79**, 3148-3151 (2016).
- Kanzaki, H., Wada, K. I., Nitoda, T. & Kawazu, K.** Novel bioactive oxazolomycin isomers produced by *Streptomyces albus* JA3453. *Biosci. Biotechnol. Biochem.* **62**, 438-442 (1998).
- Katz, L. & Baltz, R. H.** Natural product discovery: past, present and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155-176 (2016).
- Katz, L., Hover, B. M. & Brady, S. F.** Culture-independent discovery of natural products from soil metagenomes. *J. Ind. Microbiol. Biotechnol.* **43**, 129-141 (2016).
- Kaysser, L., Lutsch, L., Siebenberg, S., Wemakor, E., Kammerer, B. & Gust, B.** Identification and manipulation of the caprazamycin gene cluster lead to new simplified liponucleoside antibiotics and give insights into the biosynthetic pathway. *J. Biol. Chem.* **284**, 14987-14996 (2009).
- Kaysser, L., Siebenberg, S., Kammerer, B. and Gust, B.** Analysis of the liposidomycin gene cluster leads to the identification of new caprazamycin derivatives. *ChemBioChem* **11**, 191-196 (2010).
- Kaysser, L., Tang, X., Wemakor, E., Sedding, K., Hennig, S., Siebenberg, S. & Gust, B.** Identification of a napsamycin biosynthesis gene cluster by genome mining. *ChemBioChem* **12**, 477-487 (2011).
- Keating, T. A., Marshall, C. G. & Walsh, C. T.** Vibriobactin biosynthesis in *Vibrio cholera*: VibH is an amide synthase homologous to nonribosomal peptide synthetase condensation domains. *Biochemistry* **39**, 15513-15521 (2000).
- Keel, C., Schnider, U., Maurhofer, M., Voisard, C., Laville, J., Burger, U., Wirthner, P., Haas, D. & Défago, G.** Suppression of root diseases by *Pseudomonas fluorescens* CHA0: Importance of the bacterial secondary metabolite 2,4-diacetylphloroglucinol. *Mol. Plant Microbe Interact.* **5**, 4-13 (1992).
- Kellner, R. L. L. & Dettner, K.** Differential efficacy of toxic pederin in deterring potential arthropod predators of *Paederus* (Coleoptera: Staphylinidae) offspring. *Oecologia* **107**, 293-300.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E.** The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845-858 (2015).
- Kenney, G. E. & Rosenzweig, A. C.** Genome mining for methanobactins. *BMC Biol.* **11**, 17 (2013).

- Kikuchi, G., Motokawa, Y., Yoshida, T. & Hiraga, K.** Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. *Proc. Jpn. Acad., Ser. B.* **84**, 246-263 (2008).
- Kim, D. H., Park, J. I., Chung, S. J., Park, J. D., Park, N. K. & Han, J. H.** Cleavage of β -lactone ring by serine protease. Mechanistic implications. *Bioorg. Med. Chem.* **10**, 2553-2560 (2002).
- Kim, H. J., Graham, D. W., DiSpirito, A. A., Alterman, M. A., Galeva, N., Larive, C. K., Asunskis, D. & Sherwood, P. M. A.** Methanobactin, a copper-acquisition compound from methane-oxidising bacteria. *Science* **305**, 1612-1615 (2004).
- King, R. R., Lawrence, C. H. & Gray, J. A.** Herbicidal properties of the thaxtomin group of phytotoxins. *J. Agric. Food Chem.* **49**, 2298-2301 (2001).
- Kinscherf, T. G. & Willis, D. K.** The biosynthetic gene cluster for the β -lactam antibiotic tabtoxin in *Pseudomonas syringae*. *J. Antibiot.* **58**, 817-821 (2005).
- Kirner, S., Hammer, P. E., Hill, D. S., Altmann, A., Fischer, I., Weislo, L. J., Lanahan, M., van Pée, K. -H. and Ligon, J. M.** Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. *J. Bacteriol.* **180**, 1939-1943 (1998).
- Kitahara, M., Asano, M., Naganawa, H., Maeda, K., Hamada, M., Aoyagi, T., Umezawa, H., Iitaka, Y. & Nakamura, H.** Valilactone, an inhibitor of esterase, produced by actinomycetes. **40**, 1647-1650 (1987).
- Klahn, P. & Brönstrup, M.** Bifunctional antimicrobial conjugates and hybrid antimicrobials. *Nat. Prod. Rep.* **34**, 832-885 (2017).
- Kluge, A. F. & Petter, R. C.** Acylating drugs: redesigning natural covalent inhibitors. *Curr. Opin. Chem. Biol.* **14**, 421-427 (2010).
- Kluger, R. & Tittmann, K.** Thiamin diphosphate catalysis: enzymic and nonenzymic covalent intermediates. *Chem. Rev.* **108**, 1797-1833 (2008).
- Knaggs, A. R.** The biosynthesis of shikimate metabolites. *Nat. Prod. Rep.* **18**, 119-136 (2003).
- Ko, K., Lee, S. -H., Kim, S. -H., Jim, E. -H., Oh, K. -B., Shin, J. & Oh, D. -C.** Lajollamycins, nitro group-bearing spiro- β -lactone- γ -lactams obtained from a marine-derived *Streptomyces* sp.. *J. Nat. Prod.* **77**, 2099-2104 (2014).
- Kobayashi, J., Nakamura, M., Mori, Y., Yamashita, Y. & Kobayashi, S.** Catalytic enantio- and diastereoselective aldol reactions of glycine-derived enolate with aldehydes: an efficient approach to the asymmetric synthesis of anti- β -hydroxy- α -amino acid derivatives. *J. Am. Chem. Soc.* **126**, 9192-9193 (2004).
- Koller, W., Trail, F. & Parker, D. M.** Ebelactones inhibit cutinases produced by fungal plant pathogens. *J. Antibiot.* **43**, 734-735 (1990).
- Kondo, S., Uotani, K., Miyamoto, M., Hazato, T., Naganawa, H. & Aoyagi, T.** The structure of esterastin, an inhibitor of esterase. *J. Antibiot.* **31**, 797-800 (1978).
- Knoot, C. J., Kovaleva, E. G. & Lipscomb, J. D.** Crystal structure of CmlI, the arylamine oxygenase from the chloramphenicol biosynthetic pathway. *J. Biol. Inorg. Chem.* **21**, 589-603 (2016).

- Kridel, S. J., Axelrod, F. Rozenkrantz, N. & Smith, J. W.** Orlistat is a novel inhibitor of fatty acid synthase with antitumor activity. *Cancer Res.* **64**, 2070-2075 (2004).
- Kries, H., Wachtel, R., Pabst, A., Wanner, B., Niquille, D. & Hilvert, D.** Reprogramming nonribosomal peptide synthetases for “clickable” amino acids. *Angew. Chem. Int. Ed.* **53**, 10105-10108 (2014).
- Kries, H., Niquille, D. L. & Hilvert, D.** A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol.* **22**, 640-648 (2015)
- Kudo, F., Miyanaga, A. & Eguchi, T.** Biosynthesis of natural products containing β -amino acids. *Nat. Prod. Rep.* **31**, 1056-1073 (2014).
- Lambowitz, A. M. & Slayman, C. W.** Effect of pyrrolnitrin on electron transport and oxidative phosphorylation in mitochondria isolated from *Neurospora crassa*. *J. Bacteriol.* **112**, 1020-1022 (1972).
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & others.** Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).
- Lasken, R. S.** Genomic sequencing of uncultured microorganisms from single cells. *Nat. Rev. Microbiol.* **10**, 631-640 (2012).
- Lauinger, L., Li, J., Shostak, A., Cemel, I. A., Ha, N., Zhang, Y., Merkl, P. E., Obermeyer, S., Stankovic-Valentin, N. & others.** Thiolutin is a zinc chelator that inhibits the Prn11 and other JAMM metalloproteases. *Nat. Chem. Biol.* **13**, 709-714 (2017).
- Lautru, S., Oves-Costales, D., Pernodet, J. -L. & Challis, G. L.** MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiology* **153**, 1405-1412 (2007).
- Laville, J., Blumer, C., von Scroetter, C., Gaia, V., Keel, C. & Haas, D.** Characterization of the *hcnABC* gene cluster encoding hydrogen cyanide synthase and anaerobic regulation by ANR in the strictly aerobic biocontrol agent *Pseudomonas fluorescens* CHA0. *J. Bacteriol.* **180**, 3187-3196 (1998).
- Lederberg, J.** Pathways of discovery: infectious history. *Science* **288**, 287-293.
- Lee, J., Simurdiak, M. & Zhao, H.** Reconstitution and characterisation of aminopyrrolnitrin oxygenase, a Rieske *N*-oxygenase that catalyses unusual arylamine oxidation. *J. Biol. Chem.* **44**, 36719-36728 (2005).
- Li, Y. -M., Milne, J. C., Madison, L. L., Kolter, R. & Walsh, C. T.** From peptide to precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science* **274**, 1188-1193 (1996).
- Li, W., Ju, J., Osada, H. & Shen, B.** Utilization of the methoxymalonyl-acyl carrier protein biosynthesis locus for cloning the tautomycin biosynthetic gene cluster from *Streptomyces spiroverticillatus*. *J. Antibiot.* **188**, 4147-4151 (2006).
- Li, J. W. -H. and Vederas, J. C.** Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**, 161-165 (2009).

- Li, Q., Wang, L., Xie, Y., Wang, S., Chen, R. & Hong, B.** SsaA, a member of a novel class of transcriptional regulators, controls sansanmycin production in *Streptomyces* sp. Strain SS through a feedback mechanism. *J. Bacteriol.* **195**, 2232-2243 (2013).
- Li, R., Oliver, R. A. & Townsend, C. A.** Identification and characterisation of the sulfazecin monobactam biosynthetic gene cluster. *Cell Chem. Biol.* **24**, 24-34 (2017).
- Liberati, N. T., Urbach, J. M., Miyata, S., Lee, D. G., Drenkard, E., Wu, G., Villanueva, J., Wei, T. & Ausubel, F. M.** An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl. Acad. Sci.* **103**, 2833-2838 (2006).
- Lichstein, H. C. & Vandesand, V. F.** Violacein, an antibiotic pigment produced by *Chromobacterium violaceum*. *J. Infect. Dis.* **76**, 47-51 (1945)
- Lindeberg, M., Myers, C. R., Collmer, A. & Schneider, D. J.** Roadmap to new virulence determinants in *Pseudomonas syringae*: Insights from comparative genomics and genome organization. *Mol. Plant Microbe Interact.* **21**, 685-700 (2008).
- Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., Mueller, A., Schäberle, T. F., Hughes, D. E. & others.** A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455-459 (2015).
- Lindow, S. E., Desurmont, C., Elkins, R., McGourty, G., Clark, E. & Brandl, M. T.** Occurrence of indole-3-acetic acid-producing bacteria on pear trees and their association with fruit russet. *Phytopathology* **88**, 1149-1157 (1998).
- Lister, P. D., Wolter, D. J. & Hanson, N. D.** Antibacterial-resistant *Pseudomonas aeruginosa*: clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clin. Microbiol. Rev.* **22**, 582-610 (2009).
- Liu, L. & Shaw, P. D.** Characterization of *dapB*, a gene required by *Pseudomonas syringae* pv. *tabaci* BR2.024 for lysine and tabtoxinine- β -lactam biosynthesis. *J. Bacteriol.* **179**, 507-513 (1997).
- Liu, J. Q., Dairi, T., Itoh, N., Kataoka, M., Shimizu, S. & Yamada, H.** Gene cloning, biochemical characterisation and physiological role of a thermostable low-specificity L-threonine aldolase from *Escherichia coli*. *Eur. J. Biochem.* **255**, 220-226 (1998).
- Liu, D. -Z., Wang, F., Liao, T. -G., Tang, J. -G., Steglich, W., Zhu, H. -J. & Liu, J. -K.** Vibralactone: A lipase inhibitor with an unusual fused β -lactone produced by cultures of the basidiomycete *Boreostereum vibrans*. *Org. Lett.* **8**, 5749-5752 (2006).
- Liu, Y., Hazzard, C., Eustáquio, A. S., Reynolds, K. A. & Moore, B. S.** Biosynthesis of salinosporamides from α,β -unsaturated fatty acids: Implications for extending polyketide synthase diversity. *J. Am Chem. Soc.* **131**, 10376-10377 (2009).
- Liu, G. -S., Zhang, Y. -Q., Yuan, Y. -A. & Xu, H.** Iron(III)-catalysed intramolecular aminohydroxylation of olefins with functionalized hydroxylamines. *J. Am. Chem. Soc.* **125**, 3343-3346 (2013).

- Liu, X. & Cheng, Y. -Q.** Genome-guided discovery of diverse natural products from *Burkholderia* sp. *J. Ind. Microbiol. Biotechnol.* **41**, 275-284 (2014).
- Liu, G., Zhang, M., Chen, X., Zhang, W., Ding, W. & Zhang, Q.** Evolution of the threonine aldolases, a diverse family involved in the second pathway of glycine biosynthesis. *J. Mol. Evol.* **80**, 102-107 (2015).
- Livermore, D. M.** Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: our worst nightmare? *Clin. Infect. Dis.* **34**, 634-640 (2002).
- Llamas, M. A., Imperi, F., Visca, P. & Lamont, I. L.** Cell-surface signalling in *Pseudomonas*: stress responses, iron transport, and pathogenicity. *FEMS Microbiol. Rev.* **38**, 569-597 (2014).
- Loeschcke, A. & Thies, S.** *Pseudomonas putida* – a versatile host for the production of natural products. *Appl. Microbiol. Biotechnol.* **99**, 6197-6214 (2015).
- Loh, K. -C. & Cao, B.** Paradigm in biodegradation using *Pseudomonas putida* – A review of proteomics studies. *Enzyme Microb. Technol.* **43**, 1-12 (2008).
- López-Goñi, I., Moriyón, I. & Neilands, J. B.** Identification of 2,3-dihydroxybenzoic acid as a *Brucella abortus* siderophore. *Infect. Immun.* **60**, 4496-4503 (1992).
- Loper, J. E., Haasan, K. A., Mavrodi, D. V., Davis II, E. W., Lim, C. K., Shaffer, B. T., Elbourne, L. D. H., Stockwell, V. O., Hartney, S. L. and others.** Comparative genomics of plant-associated *Pseudomonas* spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet.* **8**, e1002784 (2012).
- Lowe, C, Pu, Y. & Vederas, J. C.** Synthesis of (+)-obafluorin, a β -lactone antibiotic. *J. Org. Chem.* **57**, 10-11 (1992).
- Lowther, J., Beattie, A. E., Langridge-Smith, P. R. R., Clarke, D. J. & Campopiano, D. J.** L-Penicillamine is a mechanism-based inhibitor of serine palmitoyltransferase by forming a pyridoxal-5'-phosphate-thiazolidine adduct. *Med. Chem. Comm.* **3**, 1003-1008 (2012).
- Lu, H., Chanco, E. & Zhao, H.** CmlI is an *N*-oxygenase in the biosynthesis of chloramphenicol. *Tetrahedron* **68**, 7651-1654 (2012).
- Lynch, M.** Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* **60**, 327-349 (2006).
- Malpartida, F. & Hopwood, D. A.** Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature* **309**, 462-464 (1984).
- Manam, R. R., Teisan, S., White, D. J., Nicholson, B., Grodberg, J., Neuteboom, S. T. C., Lam, K. S., Mosca, D. A., Lloyd, G. K. & Potts, B. C. M.** Lajollamycin, a nitro-tetraene spiro- β -lactone- γ -lactam antibiotic from the marine actinomycete *Streptomyces nodosus*. *J. Nat. Prod.* **69**, 240-243 (2005).
- Makris, T. M., Chakrabati, M., Münck, E. & Lipscomb, J. D.** A family of diiron monooxygenases catalysing amino acid beta-hydroxylation in antibiotic biosynthesis. *Proc. Natl. Acad. Sci USA* **107**, 15391-15396 (2010).

- Martins dos Santos, V. A. P., Heim, S., Moore, E. R. B., Strätz, M. & Timmis, K. N.** Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. *Environ. Microbiol.* **6**, 1264-1286 (2004).
- Masschelein, J., Mattheus, W., Gao, L. -J., Moons, P., Van Houdt, R., Uytterhoeven, B., Lamberigts, C., Lescrinier, E., Rozenski, J. & others.** A PKS/NRPS/FAS hybrid gene cluster from *Serratia plymuthica* RVH1 encoding the biosynthesis of three broad spectrum, zeamine-related antibiotics. *PLoS ONE* **8**, e54143 (2013).
- Mathee, K., Narasimhan, G., Valdes, C., Qiu, X., Matewish, J. M., Koehrsen, M., Rokas, A., Yandava, C. N., Engels, R. & others.** Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc. Natl. Acad. Sci. USA* **105**, 3100-3105 (2008).
- Mattheus, W., Masschelein, J., Gao, L. -J., Herdewijn, P., Landuyt, B., Volckaert, G. & Lavigne, R.** The kalimantacin/batumin biosynthesis operon encodes a self-resistance isoform of the FabI bacterial target. *Chem. Biol.* **17**, 1067-1071 (2010).
- Mauchline, T. H., Chedom-Fotso, D., Chandra, G., Samuels, T., Greenaway, N., Backhaus, A., McMillan, V., Canning, G., Powers, S. J. & others.** An analysis of *Pseudomonas* genomic diversity in take-all infected wheat fields reveals the lasting impact of wheat cultivars on the soil microbiota. *Environ. Microbiol.* **17**, 4764-4778 (2015).
- Mavrodi, D. V., Bonsall, R. F., Delaney, S. M., Soule, M. J., Phillips, G. & Thomashow, L. S.** Functional analysis of genes for biosynthesis of pyocyanin and phenazine-1-carboxamide from *Pseudomonas aeruginosa* PA01. *J. Bacteriol.* **183**, 6454-6465 (2001).
- Mavrodi, D. V., Blankenfeldt, W. & Thomashow, L. S.** Phenazine compounds in fluorescent *Pseudomonas* spp. biosynthesis and regulation. *Annu. Rev. Phytopathol.* **44**, 417-445 (2006).
- Mavrodi, D., Loper, J. E., Paulsen, I. T. and Thomashow, L. S.** Mobile genetic elements in the genome of the beneficial rhizobacterium *Pseudomonas fluorescens* Pf-5. *BMC Microbiol.* **9**:8 (2009).
- May, J. J., Wendrich, T. M. & Marahiel, M. A.** The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J. Biol. Chem.* **276**, 7209-7217 (2001).
- Mayol, L., Piccialli, V. & Sica, D.** Spongiolactone, an unusual β -lactone diterpene isovalerate based on a new rearranged spongiane skeleton from *Spongionella gracilis*. *Tetrahedron Lett.* **28**, 3601-3604 (1987).
- McGrath, N. A. & Raines, R. T.** Chemoselectivity in chemical biology: acyl transfer reactions with sulfur and selenium. *Acc. Chem. Res.* **44**, 752-761 (2011).
- Medema, M. H., Paavlast, Y., Nguyen, D. D., Melnik, A., Dorrestein, P. C., Takano, E. & Breitling, R.** Pep2Path: Automated mass spectrometry-guided genome mining of peptidic natural products. *PLoS Comput. Biol.* **10**, e1003822 (2014).

- Medema, M. H., Cimermancic, P., Sali, A., Takano, E. & Fischbach, M. A.** A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLoS Comp. Biol.* **10**, e1004016 (2014)
- Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J. and others.** The Minimum Information about a Biosynthetic Gene cluster (MIBiG) specification. *Nat. Chem. Biol.* **11**, 625-631 (2015).
- Medema, M. H. & Fischbach, M. A.** Computational approaches to natural product discovery. *Nat. Chem. Biol.* **11**, 639-648 (2015).
- Mehta, P. K. & Christen, P.** The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Advances in enzymology and related areas of molecular biology* **74**, 129-184 (2000).
- Melotto, M., Underwood, W., Koczan, J., Nomura, K. & He, S. Y.** Plant stomata function in innate immunity against bacterial invasion. *Cell* **126**, 969-980 (2006).
- Mercado-Blanco, J., van der Drift, K. M. G. M., Olsson, P. E., Thomas-Oates, J. E., van Loon, L. C. & Bakker, P. A. H. M.** Analysis of the *pmsCEAB* gene cluster involved in biosynthesis of salicylic acid and the siderophore pseudomonine in the biocontrol strain *Pseudomonas fluorescens* WCS374. *J. Bacteriol.* **183**, 1909-1920 (2001).
- Metelev, M., Serebryakova, M., Ghilarov, D., Zhao, Y. & Severinov, K.** Structure of microcin B-like compounds produced by *Pseudomonas syringae* and species specificity of their antibacterial action. *J. Bacteriol.* **195**, 4129-4137 (2013).
- Meyer, J. -M.** Pyoverdines: pigments, siderophores and potential taxonomic markers of fluorescent *Pseudomonas* species. *Arch. Microbiol.* **174**, 135-142 (2000).
- Meyer, S., Kehr, J. -C., Mainz, A., Dehm, D., Petras, D., Süssmuth, R. D. & Dittmann, E.** Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of NRPS. *Cell Chem. Biol.* **23**, 462-471 (2016).
- Michaux, C., Massant, J., Kerff, F., Frère, J. -M., Docquire, J. -D., Vandenberghe, I., Samyn, B., Pierrand, A., Feller, G. & others.** Crystal structure of a cold-adapted class C β -lactamase. *FEBS J.* **275**, 1687-1697 (2008).
- Miethke, M., Kraushaar, T. & Marahiel, M. A.** Uptake of xenosiderophores in *Bacillus subtilis* occurs with high affinity and enhances the folding stabilities of substrate binding proteins. *FEBS Lett.* **587**, 206-213 (2013).
- Milano, T., Paiardini, A., Grgurina, I. & Pascarella, S.** Type I pyridoxal 5'-phosphate dependent enzymatic domains embedded within multimodular nonribosomal peptide synthetase and polyketide synthase assembly lines. *BMC Struct. Biol.* **13**, 26 (2013).
- Miller, M. A., Pfeiffer, W. & Schwartz, T.** in *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 1-8 (IEE, New Orleans, LA, USA, 2010).

- Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C. & Gulick, A. M.** Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic protein architecture. *J. Biol. Chem.* **291**, 22559-22571 (2016).
- Mori, T., Takahashi, K., Kashiwabara, M., Uemura, D., Katayama, C. & Iwadare, S.** Structure of oxazolomycin, a novel β -lactone antibiotic. *Tetrahedron Lett.* **26**, 1073-1076 (1985).
- Muliandi, A., Katsuyama, Y., Sone, K., Izumikawa, M., Moriya, T., Hashimoto, J., Kozono, I., Takagi, M., Shin-ya, K. & Ohnishi, Y.** Biosynthesis of the 4-methyloxazoline-containing nonribosomal peptides, JBIR-34 and -35, in *Streptomyces* sp. Sp080513GE-23. *Chem. Biol.* **21**, 923-934 (2014).
- Müller, M., Sprenger, G. A. & Pohl, M.** C-C bond formation using ThDP-dependent lyases. *Curr. Opin. Chem. Biol.* **17**, 261-270 (2013).
- Murphy, C. D., O'Hagan, D. & Schaffrath, C.** Identification of a PLP-dependent threonine transaldolase: a novel enzyme involved in 4-fluorothreonine biosynthesis in *Streptomyces cattleya*. *Angew. Chem. Int. Ed.* **40**, 4479-4481 (2001).
- Nakamura, M. H., Honma, M., Kamada, T., Ohno, T., Kunimoto, S., Ikeda, Y., Kondo, S. & Takeuchi, T.** Inhibitory effect of curromycin A and B on human immunodeficiency virus replication. *J. Antibiot.* **47**, 616-618 (1994).
- Nardini, M. & Dijkstra, B. W.** α/β Hydrolase fold enzymes: the family keeps growing. *Curr. Opin. Struct. Biol.* **9**, 732-737 (1999).
- Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V. A. P., Fouts, D. E., Gill, S. R., Pop, M. & others.** Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 799-808 (2002).
- Newman, D. J., Cragg, G. M. & Snader, K. M.** The influence of natural products upon drug discovery. *J. Nat. Prod.* **17**, 215-234 (2000).
- Newman, D. J. & Cragg, G. M.** Natural products as the sources of new drugs over the last 25 years. *J. Nat. Prod.* **70**, 461-477 (2007).
- Nguyen, D. D., Melnik, A. V., Koyama, N., Lu, X., Schorn, M., Fang, J., Aguinaldo, K., Lincecum, T. L., Ghequire, M. G. K., & others.** Indexing the *Pseudomonas* specialized metabolome enabled the discovery of Poaeamide B and the banamides. *Nat. Microbiol.* **2**, 16197 (2016).
- Nichols, B. P., Seibold, A. M. & Doktor, S. Z.** *para*-Aminobenzoate biosynthesis from chorismate occurs in two steps. *J. Biol. Chem.* **264**, 8597-8601 (1989).
- Nobary, S. G. & Jensen, S. E.** A comparison of the clavam biosynthetic gene clusters in *Streptomyces antibioticus* Tü1718 and *Streptomyces clavuligerus*. *Can. J. Microbiol.* **58**, 413-425 (2012).
- Noike, M., Matsui, T., Ooya, K., Sasaki, I., Ohtaki, S., Hamano, Y., Maruyama, C., Ishikawa, J., Satoh, Y. & others.** A peptide ligase and the ribosome cooperate to synthesize the peptide peganomycin. *Nat. Chem. Biol.* **11**, 71-78 (2014).
- Nonaka, Y., Ohtaki, H., Ohtsuka, E., Kocho, T., Fukuda, T., Takeuchi, T. & Aoyagi, T.** Effects of ebelactone B, a lipase inhibitor, on intestinal absorption in the rat. *J. Enzyme Inhib.* **10**, 57-63 (1996).

- Noordman, W. H. & Janssen, D. B.** Rhamnolipid stimulates uptake of hydrophobic compounds by *Pseudomonas aeruginosa*. **68**, 4502-4508 (2002).
- Notredame, C., Higgins, D. G. & Heringa, J.** T-Coffee: a novel method for fast accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205-217 (2000).
- Nowak-Thompson, B. & Gould, S. J.** Biosynthesis of the pseudobactin chromophore from tyrosine. *Tetrahedron* **50**, 9865-9872 (1994).
- Nowak-Thompson, B., Gould, S. J., Krau, J. & Loper, J. E.** Production of 2,4-diacetylphloroglucinol by the biocontrol agent *Pseudomonas fluorescens* Pf-5. *Can. J. Microbiol.* **40**, 1064-1066 (1994).
- Nowak-Thompson, B., Gould, S. J. & Loper, J. E.** Identification and sequence analysis of the genes encoding a polyketide synthase required for pyoluteorin biosynthesis in *Pseudomonas fluorescens* Pf-5. *Gene* **204**, 17-24 (1997).
- Nowak-Thompson, B., Chaney, N., Wing, J. S., Gould, S. J. & Loper, J. E.** Characterization of the pyoluteorin biosynthetic gene cluster of *Pseudomonas fluorescens* Pf-5. *J. Bacteriol.* **181**, 2166-2174 (1999).
- Nowak-Thompson, B., Hammer, P. E., Hill, D. S., Stafford, J., Torkewitz, N., Gaffney, T. D., Lam, S. T., Molnár, I. & Ligon, J. M.** 2,5-Dialkylresorcinol biosynthesis in *Pseudomonas aurantiaca*: Novel head-to-head condensation of two fatty acid-derived precursors. *J. Bacteriol.* **185**, 860-869 (2003).
- Ogura, M., Nakayama, H., Furihata, K., Shimazu, A., Seto, H. & Otake, N.** Structure of a new antibiotic curromycin A produced by a genetically modified strain of *Streptomyces hygrosopicus*, a polyether antibiotic producing organism. *J. Antibiot.* **38**, 669-673 (1985).
- Okegbe, C., Fields, B. L., Cole, S. J., Beierschmitt, C., Morgan, C. J., Price-Whelan, A., Stewart, R. C., Lee, V. T. & Dietrich, L. E. P.** Electron-shuttling antibiotics structure bacterial communities by modulating cellular levels of c-di-GMP. *Proc. Natl. Acad. Sci. USA* **114**, E5236-E5245 (2017).
- Olano, C., Wilkinson, B., Sánchez, C., Moss, S. J., Sheridan, R., Math, V., Weston, A. J., Braña, A. F., Martin, C. J. & others.** Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: cluster analysis and assignment of functions. *Chem. Biol.* **11**, 87-97 (2004).
- Oldfield, E. & Lin, F. -Y.** Terpene biosynthesis: Modularity rules. *Angew. Chem. Int. Ed.* **51**, 1124-1137 (2012).
- Oliynyk, M., Samborsky, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F. and Leadlay, P. F.** Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol.* **25**, 447-453 (2007).
- Ollagnier-de Choudens, S., Loiseau, L., Sanakis, Y., Barras, F. & Fontecave, M.** Quinolinate synthetase, an iron-sulfur enzyme in NAD biosynthesis. *FEBS Lett.* **579**, 3737-3743 (2005).
- Ooya, K., Ogasawara, Y., Noike, M. & Dairi, T.** Identification and analysis of the resorcinomycin biosynthetic gene cluster. *Biosci. Biotechnol. Biochem.* **79**, 1833-1837 (2015).

- O'Rourke, S., Widdick, D. & Bibb, M. J.** A novel mechanism of immunity controls the onset of cinnamycin biosynthesis in *Streptomyces cinnamoneus* DSM 40646. *J. Ind. Microbiol. Biotechnol.* **44**, 563-572 (2017).
- Ortholand, J. -Y. & Ganesan, A.** Natural products and combinatorial chemistry: back to the future. *Curr. Opin. Chem. Biol.* **8**, 271-280 (2004).
- O'Shea, R. & Moser, H. E.** Physicochemical properties of antibacterial compounds: Implications for drug discovery. *J. Med. Chem.* **51**, 2871-2878 (2008).
- Osborn, A.** Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* **26**, 449-457 (2010).
- Ostrowska, H., Kalinowska, J., Chabielska, E., Stankiewicz, A., Kruszewski, K., Buczko, W.** Ebelactone B, an inhibitor of extracellular cathepsin A-type enzyme, suppresses platelet aggregation *ex vivo* in renovascular hypertensive rats. *J. Cardiovasc. Pharmacol.* **45**, 348-353 (2005).
- Otani, T., Yoshida, K. I., Kubota, H., Kawai, S., Ito, S., Hori, H., Ishiyama, T. & Oki, T.** Novel triene- β -lactone antibiotic, oxazolomycin derivative and its isomer, produced by *Streptomyces* sp. KSM-2690. *J. Antibiot.* **53**, 1397-1400 (2000).
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. & Stevens, R.** The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, D206-D214 (2014).
- Paiardini, A. Contestabile, R., D'Aguanno, S., Pascarella, S. & Bossa, F.** Threonine aldolase and alanine racemase: novel examples of convergent evolution in the superfamily of vitamin B6-dependent enzymes. *Biochim. Biophys. Acta* **1647**, 214-219 (2003).
- Parry, R., Nishino, S. & Spain, J.** Naturally-occurring nitro compounds. *Nat. Prod. Rep.* **28**, 152-167 (2011).
- Parsons, J. F., Calabrese, K., Eisenstein, E. & Ladner, J. E.** Structure of the phenazine biosynthesis enzyme PhzG. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **60**, 2110-2113 (2004).
- Parsons, J. F., Greenhagen, B. T., Shi, K., Calabrese, K., Robinson, H. & Ladner, J. E.** Structural and functional analysis of the pyocyanin biosynthetic protein PhzM from *Pseudomonas aeruginosa*. *Biochemistry* **20**, 1821-1828 (2007).
- Patten, C. L. & Glick, B. R.** Bacterial biosynthesis of indole-3-acetic acid. *Can. J. Microbiol.* **42**, 207-220 (1996).
- Patten, C. L. & Glick, B. R.** Role of *Pseudomonas putida* indoleacetic acid in development of the host plant root system. *Appl. Environ. Microbiol.* **68**, 3795-3801 (2002).
- Paulsen, I. T., Brown, M. H. & Skurray, R. A.** Proton-dependent multidrug efflux systems. *Microbiol. Rev.* **60**, 575-608 (1996).

- Paulsen, I. T., Press, C. M., Ravel, J., Kobayashi, D. Y., Myers, G. S. A., Mavrodi, D. V., DeBoy, R. T., Seshadri, R., Ren, Q. & others.** Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat. Biotechnol.* **23**, 873-878 (2005).
- Peet, R. C., Lindgren, P. B., Willis, D. K. & Panopoulos, N. J.** Identification and cloning of genes involved in phaseolotoxin production by *Pseudomonas syringae* pv. *phaseolicola*. *J. Bacteriol.* **166**, 1096-1105 (1986).
- Pence, H. E., & Williams, A.** ChemSpider: an online chemical information resource. *J. Chem. Educ.* **87**, 1123-1124 (2010).
- Peng, C., Pu, J. -Y., Song, L. -Q., Jian, X. -H., Tang, M. -C. & Tang, G. -L.** Hijacking a hydroxyethyl unit from a central metabolic ketose into a nonribosomal peptide assembly line. *Proc. Natl. Acad. Sci. USA* **109**, 8540-8545 (2012).
- Philippon, A., Labia, R. & Jacoby, G.** Extended spectrum beta-lactamases. *Antimicrob. Agents Chemother.* **28**, 302-307 (1985).
- Piel, J.** A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl. Acad. Sci. USA* **99**, 14002-14007 (2002).
- Piel, J., Hui, D., Wen, G., Butzke, D., Platzner, M., Fusetani, N. & Matsunaga, S.** Antitumour polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Entotheonella swinhoei*. *Proc. Natl. Acad. Sci. USA* **101**, 16222-16227 (2004).
- Pierson III, L. S., Gaffney, T., Lam, S. & Gong, F.** Molecular analysis of genes encoding phenazine biosynthesis in the biological control bacterium *Pseudomonas aureofaciens* 30-84. *FEMS Microbiol. Lett.* **134**, 299-307 (1995).
- Pittard, J., Camakaris, H. & Yang, J.** The TyrR regulon. *Mol. Microbiol.* **55**, 16-26 (2005).
- Podust, L. M. & Sherman, D. H.** Diversity of P450 enzymes in biosynthesis of natural products. *Nat. Prod. Rep.* **29**, 1251-1266 (2012).
- Pohl, M., Sprenger, G. A. & Müller, M.** A new perspective on thiamine catalysts. *Curr. Opin. Biotechnol.* **15**, 335-342 (2004).
- Pramanik, A. & Braun, V.** Albomycin uptake via a ferric hydroxamate transport system of *Streptococcus pneumoniae* R6. *J. Bacteriol.* **188**, 3878-3886 (2006).
- Price-Whelan, A., Dietrich, L. E. P. & Newman, D. K.** Rethinking 'secondary' metabolism: physiological roles for phenazine antibiotics. *Nat. Chem. Biol.* **2**, 71-78 (2006).
- Proschak, A., Zhou, Q., Schöner, T., Thanwisai, A., Kresovic, D., Dowling, A., French-Constant, R., Proschak, E. & Bode, H. B.** Biosynthesis of the insecticidal xenoclyoins in *Xenorhabdus bovienii*. *ChemBioChem* **15**, 369-372 (2014).
- Pu, Y., Lowe, C., Sailer, M. & Vederas, J.** Synthesis, stability, and antimicrobial activity of (+)-obafluorin and related β -lactone antibiotics. *J. Org. Chem.* **59**, 3642-3655 (1994).

- Putzer, H., Brakhage, A. A. & Grunberg-Manago, M.** Independent genes for two threonyl-tRNA synthetases in *Bacillus subtilis*. *J. Bacteriol.* **172**, 4593-4602 (1990).
- Putzer, H., Gendron, N. & Grunberg-Manago, M.** Co-ordinate expression of the two threonyl-tRNA synthetase genes in *Bacillus subtilis*: control by transcriptional antitermination involving a conserved regulatory sequence. *EMBO J.* **11**, 3117-3127 (1992).
- Qin, N., Callahan, S. M., Dunlap, P. V. & Stevens A. M.** Analysis of LuxR regulon gene expression during quorum sensing in *Vibrio fischeri*. *J. Bacteriol.* **189**, 4127-4134 (2007).
- Qin, Z., Munnoch, J. T., Devine, R., Holmes, N. A., Seipke, R. F., Wilkinson, K. A., Wilkinson, B. & Hutchings, M. I.** Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African *Tetraponera* plant-ants. *Chem. Sci.* **8**, 3218-3227 (2017).
- Quadri, L. E. N., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T.** Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-confirming siderophore mycobactin. *Chem. Biol.* **5**, 631-645 (1998).
- Quadri, L. E. N., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P. & Walsh, C. T.** Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585-1595 (1998).
- Raaijmakers, J. M. & Weller, D. M.** Natural plant protection by 2,4-diacetylphlorogluciol-producing *Pseudomonas* spp. in Take-All decline soils. *Mol. Plant Microbe Interact.* **11**, 144-152 (1998).
- Raaijmakers, J. M., de Bruijn, I. & de Kock, J. D.** Cyclic lipopeptide production by plant-associated *Pseudomonas* spp.: Diversity, activity, biosynthesis, and regulation. *Mol. Plant Microbe Interact.* **7**, 699-710 (2006).
- Rachid, S., Huo, L., Herrmann, J., Stadler, M., Köpcke, B., Bitzer, J. & Müller, R.** Mining the cinnabaramide biosynthetic pathway to generate novel proteasome inhibitors. *ChemBioChem* **12**, 922-931 (2011).
- Rao, M. N., Holkar, A. G. & Ayyangar, N. R.** A novel transacylation method for the synthesis of α -N-acyl- β -lactones; application to (\pm)-diacetylabafluorin and (+)-SQ 26,517. *J. Chem. Soc. Chem. Comm.* 1007-1008 (1991).
- Rascher, A., Hu, A. Z., Viswanathan, N., Schirmer, A., Reid, R., Nierman, W. C., Lewis, M. & Hutchinson, C. R.** Cloning and characterization of a gene cluster for geldanamycin production in *Streptomyces hygrosopicus* NRRL 3602. *FEMS Microbiol. Lett.* **218**, 223-230 (2003).
- Ravel, J. & Cornelis, P.** Genomics of pyoverdine-mediated iron uptake in pseudomonads. *Trends Microbiol.* **11**, 195-200 (2003).
- Reimann, C., Serino, L., Beyeler, M. & Haas, D.** Dihydroaeruginosic acid synthetase and pyochelin synthetase, products of the *pchEF* genes, are induced by extracellular pyochelin in *Pseudomonas aeruginosa*. *Microbiology* **144**, 3135-3148 (1998).

- Rezzonico, F., Zala, M., Keel, C. Duffy, B., Moëgne-Loccoz, Y. & Défago, G.** Is the ability of biocontrol fluorescent pseudomonads to produce the antifungal metabolite 2,4-diacetylphloroglucinol really synonymous with higher plant protection? *New Phytologist* **173**, 861-872 (2007).
- Rice, L. B.** Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. *J. Infect. Dis.* **197**, 1079-1081 (2008).
- Rodrigues, T., Reker, D., Schneider, P. & Schneider, G.** Counting on natural products for drug design. *Nat. Chem.* **8**, 531-541 (2016).
- Ross, J. I., Eady, E. A., Cove, J. H., Cunliffe, W. J., Baumberg, S. & Wootton, J. C.** Inducible erythromycin resistance in staphylococci is encoded by a member of the ATP-binding transport super gene family. *Mol. Microbiol.* **4**, 1207-1214 (1990).
- Roth, P., Hädener, A. & Tamm, C.** Further studies on the biosynthesis of tabtoxin (wildfire toxin): Incorporation of [2,3-¹³C₂]pyruvate into the β-lactam moiety.
- Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C. & Kohlbacher, O.** NRPSpredictor2 – a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362-367 (2011). *Helv. Chim. Acta* **73**, 476-482 (1990).
- Roux, M. V., Jiménez, P., Dávalos, J. Z., Castaño, O., Molina, M. T., Notario, R. M., Herreros, M. & Abboud, J. -L. M.** The first direct experimental determination of strain in neutral and protonated 2-azetidinone. *J. Am. Chem. Soc.* **118**, 12735-12737 (1996).
- Ruiz, N., Falcone, B., Kahne, D. & Silhavy, T. J.** Chemical conditionality: A genetic strategy to probe organelle assembly. *Cell* **121**, 307-317 (2005).
- Ruiz, N., Kahne, D. & Silhavy, T. J.** Advances in understanding bacterial outer-membrane biogenesis. *Nat. Rev. Microbiol.* **4**, 57-66 (2006).
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B.** Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945 (2000).
- Rutledge, P. J. & Challis, G. L.** Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13**, 509-523 (2015).
- Sansinenea, E. & Ortiz, A.** Secondary metabolites of soil *Bacillus* spp. *Biotechnol. Lett.* **33**, 1523-1538 (2011).
- Sankaranarayanan, R., Dock-Bregeon, A. -C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B. & Moras, D.** The structure of threonyl-tRNA synthetase-tRNA^{thr} complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **97**, 371-381 (1999).
- Sankaranarayanan, R., Dock-Bregeon, A. -C., Rees, B., Bovee, M., Caillet, J., Romby, P., Francklyn, C. S. & Moras, D.** Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. *Nat. Struct. Biol.* **7**, 461-465 (2000).
- Sattely, E. S. & Walsh, C. T.** A latent oxazoline electrophile for N-O-C bond formation in pseudomonine biosynthesis. *J. Am. Chem. Soc.* **17**, 12282-12284 (2008).

- Schaffer, J. E., Reck, M. R., Prasad, N. K. & Wencewicz, T. A.** β -Lactone formation during product release from a nonribosomal peptide synthetase. *Nat. Chem. Biol.* **13**, 737-744 (2017).
- Scheible, W. -R., Fry, B., Kochevenko, A., Schindelasch, D., Zimmerli, L., Somerville, S., Loria, R. & Somerville, C. R.** An arabidopsis mutant resistant to thaxtomin A, a cellulose synthesis inhibitor from *Streptomyces* species. *Plant Cell* **15**, 1781-1794 (2003).
- Schertzer, J. W., Brown, S. A. & Whiteley, M.** Oxygen levels rapidly modulate *Pseudomonas aeruginosa* social behaviours via substrate limitation of PqsH. *Mol. Microbiol.* **77**, 1527-1538 (2010).
- Schevchenko, A., Wilm, M., Vorm, O. & Mann, M.** Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **68**, 850-858 (1996).
- Schirch, V., Hopkins, S., Villar, E. & Angelaccio, S.** Serine hydroxymethyltransferase from *Escherichia coli*: purification and properties. *J. Bacteriol.* **163**, 1-7 (1985).
- Schneider, C.** Catalytic, enantioselective syntheses of β -lactones – versatile synthetic building blocks in organic chemistry. *Angew. Chem. Int. Ed.* **41**, 744-746 (2002).
- Scholz-Schroeder, B. K., Soule, J. D. & Gross, D. C.** The *sypA*, *sypB*, and *sypC* synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. *syringae* B301D. *Mol. Plant Microbe Interact.* **16**, 271-280 (2003).
- Schöner, T. A., Gassel, S., Osawa, A., Tobias, N. J., Okuno, Y., Sakakibara, Y., Shindo, K., Sandmann, G. & Bode, H. B.** Aryl polyenes, a highly abundant class of bacterial natural products, are functionally related to antioxidative carotenoids. *ChemBioChem* **17**, 247-253 (2016).
- Schneider, G., Käck, H. & Lindqvist, Y.** The manifold of vitamin B₆ dependent enzymes. *Structure* **8**, R1-R6 (2000).
- Schuhr, C. A., Eisenreich, W., Goese, M., Stohler, P., Weber, W., Kupfer, E. & Bacher, A.** Biosynthetic precursors of the lipase inhibitor lipstatin. *J. Org. Chem.* **67**, 2257-2262 (2002).
- Schwyn, B. & Neilands, J. B.** Universal chemical assay for the detection and determination of siderophores. *Anal. Biochem.* **160**, 47-56 (1987).
- Scott, T. A., Heine, D., Qin, Z. & Wilkinson, B.** An L-threonine transaldolase is required for L-threo- β -hydroxy- α -amino acid assembly during obafluorin biosynthesis. *Nat. Commun.* **8**, 15935 (2017).
- Seiple, I. B., Mercer, J. A. M., Sussman, R. J., Zhang, Z. & Myers, A. G.** Stereocontrolled synthesis of syn- β -hydroxy- α -amino acids by direct aldolization of pseudoephedrine glycinamide. *Angew. Chem. Int. Ed. Engl.* **53**, 4642-4647 (2014).
- Sela, I., Wolf, Y. I. & Koonin, E. V.** Theory of prokaryotic genome evolution. *Proc. Natl. Acad. Sci. USA* **113**, 11399-11407 (2016).

- Serino, L., Reimmann, C., Visca, P., Beyeler, M., Chiesa, V. D. & Haas, D.** Biosynthesis of pyochelin and dihydroaeruginosic acid requires the iron-regulated *pchDCBA* operon in *Pseudomonas aeruginosa*. *J. Bacteriol.* **179**, 248-257 (1997).
- Severinov, K., Semenova, E., Kazakov, A., Kazakov, T. & Gelfand, M. S.** Low-molecular-weight post-translationally modified microcins. *Mol. Microbiol.* **65**, 1380-1394 (2007).
- Schaffrath, C., Cobb, S. L. & O'Hagan, D.** Cell-free biosynthesis of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. *Angew. Chem. Int. Ed.* **114**, 4069-4071 (2002).
- Shah, I. M., Lees, K. R., Pien, C. P. & Elliott, P. J.** Early clinical experience with the novel proteasome inhibitor PS-519. *Br. J. Clin. Pharmacol.* **54**, 269-276 (2002).
- Shanahan, P., Glennon, J. D., Crowley, J. J., Donnelly, D. F. & O'Gara, F.** Liquid-chromatographic assay of microbially derived phloroglucinol antibiotics for establishing the biosynthetic route to production, and the factors affecting their regulation. *Anal. Chim. Acta* **272**, 271-277 (1993).
- Shimidzu, Y., Ogata, H. & Goto, S.** Type III polyketide synthases: Functional classification and phylogenomics. *ChemBioChem* **18**, 50-65 (2017).
- Shin, H. S. & Rogers, P. L.** Biotransformation of benzaldehyde to l-phenylacetylcarbinol, and intermediate in l-ephedrine production, by immobilized *Candida utilis*. *Appl. Microbiol. Biotechnol.* **44**, 7-14 (1995).
- Shiomi, K., Arai, N., Shinose, M., Takahashi, Y., Yoshida, H., Iwabuchi, J., Tanaka, Y. & Omura, S.** New antibiotics phthoxazolins B, C and D produced by *Streptomyces* sp. KO-7888. *J. Antibiot.* **48**, 714-719 (1995).
- Silakowski, B., Kunze, B., Nordsiek, G., Blöcker, H., Höfle, G. & Müller, R.** The myxochelin iron transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15. *Eur. J. Biochem.* **267**, 6476-6485 (2000).
- Silby, M. W., Cerdeño-Tárraga, A. M., Vernikos, G. S., Giddens, S. R., Jackson, R. W., Preston, G. M., Zhang, X. -X., Moon, C. D., Gehrig, S. M. & others.** Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biol.* **10**, R51 (2009).
- Silby, M. W., Winstanley, C., Godfrey, S. A. C., Levy, S. B. & Jackson, R. W.** *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol. Rev.* **35**, 652-680 (2011).
- Simon, R., Priefer, U. & Pühler, A.** A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. *Nat. Biotechnol.* **1**, 784-791 (1983).
- Skaff, D. A., Ramyar, K. X., McWhorter, W. J., Barta, M. L., Geisbrecht, B. V. & Miziorko, H. M.** The biochemical and structural basis for inhibition of *Enterococcus faecalis* HMG-CoA synthase, *mvaS*, by hymeglusin. *Biochemistry* **51**, 4713-4722 (2012).

- Slock, J., Stahly, D. P., Han, C. -Y., Six, E. W. & Crawford, I. P.** An apparent *Bacillus subtilis* folic acid biosynthetic operon containing *pab*, an amphibolic *trpG* gene, a third gene required for synthesis of *para*-aminobenzoic acid and the dihydropteroate synthetase gene. *J. Bacteriol.* **172**, 7211-7226 (1990).
- Smidt, M. & Kosuge, T.** The role of indole-3-acetic acid accumulation by alpha methyl tryptophan-resistant mutants of *Pseudomonas savastanoi* in gall formation of oleanders. *Physiol. Plant Pathol.* **13**, 203-214 (1978).
- Soucy, F., Grenier, L., Behnke, M. L., Destree, A. T., McCormack, T. A., Adams, J. & Phamondon, L.** A novel and efficient synthesis of a highly active analogue of clasto-lactacystin β -lactone. *J. Am. Chem. Soc.* **121**, 9967-9976 (1999).
- Spaepen, S., Vanderleyden, J. & Remans, R.** Indole-3-acetic acid in microbial and microorganism-plant signalling. *FEMS Microbiol. Rev.* **31**, 425-448 (2007).
- Spencer, D. H., Kas, A., Smith, E. E., Raymond, C. K. Sims, E. H., Hastings, M., Burns, J. L., Kaul, R. & Olsen, M. V.** Whole-genome sequence variation among multiple isolates of *Pseudomonas aeruginosa*. *J. Bacteriol.* **185**, 1316-1325 (2003).
- Stachelhaus, T., Mootz, H. D. & Marahiel, M. A.** The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493-505 (1999).
- Stadler, H., Oesterhelt, G. & Borgström, B.** Tetrahydrolipstatin: degradation products produced by human carboxyl-ester lipase. *Helv. Chim. Acta* **75**, 1593-1603 (1992).
- Stamatakis, A.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
- Staunton, J. & Weissman, K. J.** Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380-416 (2001).
- Stefanska, A. L., Fulston, M., Houge-Frydrych, C. S. V., Jones, J. J. & Warr, S. R.** A potent tRNA synthetase inhibitor SB-217452 isolated from a *Streptomyces* species. *J. Antibiot.* **53**, 1346-1353 (2000).
- Steinreiber, J., Fesko, K., Mayer, C., Reisinger, C., Schürmann, M. & Griengl, H.** Synthesis of γ -halogenated and long-chain β -hydroxy- α -amino acids and 2-amino-1,3-diols using threonine aldolases. *Tetrahedron* **63**, 8088-8093 (2007).
- Steinreiber, J., Fesko, K., Reisinger, C., Schürmann, M., van Assema, F., Wolberg, M., Mink, D. & Griengl, H.** Threonine aldolases – an emerging tool for organic synthesis. *Tetrahedron* **63**, 918-926 (2007).
- Stewart, P. S. & Costerton, J. W.** Antibiotic resistance of bacteria in biofilms. *Lancet* **358**, 135-138 (2001).
- Stothard, P.** The sequence manipulation suite: JavaScript programs for analysing and formatting protein and DNA sequences. *Biotechniques* **28**, 1102-1104 (2000).
- Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S. L., Hufnagle, W. O., Kowalik, D. J. & others.** Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **31**, 959-964 (2000).

- Su, L., Lv, M., Kyeremeh, K., Deng, Z., Deng, H. & Yu, Y.** A ThDP-dependent enzymatic carbonylation reaction involved in neocarazostatin A tricyclic carbazole formation. *Org. Biomol. Chem.* **14**, 8679-8684 (2016).
- Süssmuth, R. D. & Mainz, A.** Nonribosomal peptide synthesis – principles and prospects. *Angew. Chem. Int. Ed.* **56** 3770-3821 (2017).
- Takahashi, K., Kawabata, M. & Uemura, D.** Structure of neooxazolomycin, an antitumor antibiotic. *Tetrahedron Lett.* **26**, 1077-1078 (1985).
- Takano, E., White, J., Thompson, C. J. and Bibb, M. J.** Construction of thioestrepton-inducible, high-copy-number expression vectors for use in *Streptomyces* spp. *Gene* **166**, 133-137 (1995).
- Télez, C. M., Gaus, K. P., Graham, D. W., Arnold, R. G. & Guzman, R. Z.** Isolation of copper biochelates from *Methylosinus trichosporium* OB3b and soluble methane monooxygenase mutants. *Appl. Environ. Microbiol.* **64**, 1115-1122 (1998).
- Thaker, M. N., Wang, W., Spanogiannopoulos, P., Waglechner, N., King, A. M., Medina, R. & Wright, G. D.** Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat. Biotechnol.* **31**, 922-927 (2013).
- Thirlway, J., Lewis, R., Al Nakeeb, M. Styles, M., Struck, A. -W., Smith, C. and Micklefield, J.** Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew. Chem. Int. Ed.* **51**, 7181-7184 (2012).
- Thomashow, L. S. and Weller, D. M.** Role of a phenazine antibiotic from *Pseudomonas fluorescens* in biological control of *Gaeumannomyces graminis* var. *tritici*. **170**, 3499-3508 (1988).
- Tobias, N. J., Heinrich, A. K., Eresmann, H., Wright, P. R., Neubacher, N., Backofen, R. & Bode, H. B.** *Photorhabdus*-nematode symbiosis is dependent on *hfq*-mediated regulation of secondary metabolites. *Environ. Microbiol.* **19**, 119-129 (2017).
- Torgerson, J. S., Hauptman, J., Boldrin, M. N. & Sjöström, L.** XENicalin the prevention of diabetes in obese subjects (XENDOS) study: a randomized study of orlistat as an adjunct to lifestyle changes for the prevention of type 2 diabetes in obese patients. *Diabetes Care* **27**, 155-161 (2004).
- Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A. & Walsh, C. T.** Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**, 215-218 (2000).
- Traxler, M. F. & Kolter, R.** Natural products in soil microbe interactions and evolution. *Nat. Prod. Rep.* **32**, 956-970 (2015).
- Tripathi, R. K. and Gottlieb, D.** Mechanism of action of the antifungal antibiotic pyrrolnitrin. *J. Bacteriol.* **100**, 310-318 (1969).
- Trost, B. M. & Miege, F.** Development of ProPhenol ligands for the diastereo- and enantioselective synthesis of β -hydroxy- α -amino esters. *J. Am. Chem. Soc.* **136**, 3016-3019 (2014).
- Tsukiura, H., Miyaki, T., Kitazima, H., Okanishi, M., Ohmori, T., Kawaguchi, H. & Koshiyama, H.** Danomycin, a new antibiotic. *J. Antibiot.* **17**, 39-47 (1964).

- Tymiak, A. A., Culver, A. A., Malley, M. F. & Gougoutas, J. Z.** Structure of obafluorin. An antibacterial β -lactone from *Pseudomonas fluorescens*. *J. Org. Chem.* **50**, 5491-5495 (1985).
- Udwarý, D. W., Zeigler, L., Asoklar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R. & Moore, B. S.** Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinospora tropica*. *Proc. Natl. Acad. Sci. USA* **104**, 10376-10381 (2007).
- Ueoka, R., Uria, A. R., Reiter, S., Mori, T., Karbaum, P., Peters, E. E., Helfrich, E. J. N., Morinaka, B. I., Gugger, M. & others.** Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nat. Chem. Biol.* **11**, 705-712 (2015).
- Unkefer, C. J., London, R. E., Durbin, R. D., Uchytýl, T. F. & Langston-Unkefer, P. J.** The biosynthesis of tabtoxinine- β -lactam. *J. Biol. Chem.* **262**, 4994-4999 (1987).
- Uppalapati, S. R., Ishiga, Y., Wangdi, T., Kunkel, B. N., Anand, A., Mysore, K. S. & Bender, C. L.** The phytotoxin coronatine contributes to pathogen fitness and is required for suppression of salicylic acid accumulation in tomato inoculated with *Pseudomonas syringae* pv. *tomato* DC3000. *Mol. Plant Microbe Interact.* **20**, 955-965 (2007).
- Umezawa, H., Aoyagi, T., Suda, H., Hamada, M., & Takeuchi, T.** Bestatin, an inhibitor of aminopeptidase-B, produced by actinomycetes. *J. Antibiot.* **29**, 97-99 (1976).
- Umezawa, H., Aoyagi, T., Hazato, T., Uotani, K., Kojima, F., Hamada, M. & Takeuchi, T.** Esterastin, an inhibitor of esterase, produced by actinomycetes. *J. Antibiot.* **31**, 639-641 (1978).
- Umezawa, H., Aoyagi, T., Uotani, K., Hamada, M., Takeuchi, T. & Takahashi, S.** Ebelactone, an inhibitor of esterase, produced by actinomycetes. *J. Antibiot.* **33**, 1594-1596 (1980).
- Van Pée, K. -H. & Ligon, J. M.** Biosynthesis of pyrrolnitrin and other phenylpyrrole derivatives in bacteria. *Nat. Prod. Rep.* **17**, 157-164 (2000).
- Vertesy, L., Aretz, W., Fehlhaber, H. W. & Kogler, H.** Salmycin A-D, antibiotics from *Streptomyces violaceus*, DSM-8286, having a siderophore aminoglycoside structure. *Helv. Chim. Acta* **78**, 46-60 (1995).
- Visca, P., Imperi, F. and Lamont I. L.** Pyoverdine siderophores: from biogenesis to biosignificance. *Trends Microbiol.* **15**, 22-30 (2006).
- Vodovar, N., Vallenet, D., Cruveiller, S., Rouy, Z., Barbe, V., Acosta, C., Cattolico, L., Jubin, C., Lajus, A. & others.** Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*. *Nat. Biotechnol.* **24**, 673-679 (2006).
- Vogel, C. & Pleiss, J.** The modular structure of ThDP-dependent enzymes. *Proteins* **82**, 2523-2537 (2014).

- Voisard, C., Bull. C.T., Keel, C., Laville, J., Maurhofer, M., Schnider, U., Défago, G. & Haas, D.** Biocontrol of root diseases by *Pseudomonas fluorescens* CHA0: current concepts and experimental approaches. In *Molecular Ecology of Rhizosphere Microorganisms* (O'Gara, F., Dowling, D. N. and Boesten, B., eds). Weinheim, Germany: VCH, pp. 67-89 (1994).
- Walsh, C. T., Liu, J., Rusnak, F. & Sakaitani, M.** Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway. *Chem. Rev.* **90**, 1105-1129 (1990).
- Walsh, C. T., Fisher, C. L., Park, I. -S., Prahland, M. & Wu, Z.** Bacterial resistance to vancomycin: five genes and one missing hydrogen tell the story. *Chem. Biol.* **3**, 21-28 (1996).
- Walsh, C. T.** Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**, 775-781 (2000).
- Wandersman, C. & Delepelaire, P.** Bacterial iron homeostasis: From siderophores to hemophores. *Annu. Rev. Microbiol.* **58**, 611-647 (2004).
- Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapon, C. A. & others.** Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828-837 (2016).
- Watrous, J. D. & Dorrestein, P. C.** Imaging mass spectrometry in microbiology. *Nat. Rev. Microbiol.* **9**, 683-694 (2011).
- Watrous, J. D., Roach, P., Alexanderov, T., Heath, B. S. Yang, J. Y., Kersten, R. D., van der Voort, M., Pogliano, K., Gross, H. & others.** Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. USA* **109**, E1743-E1752 (2012).
- Weibel, E. K., Hadváry, P., Hochuli, E., Kupfer, E. & Lengsfeld, H.** Lipstatin, an inhibitor of pancreatic lipase, produced by *Streptomyces toxytricini*. *J. Antibiot.* **40**, 1081-1085 (1987).
- Weinandy, F., Lorenz-Baath, K., Korotkov, V. S., Böttcher, T., Sethi, S., Chakraborty, T. & Sieber, S. A.** A β -lactone-based antivirulence drug ameliorates *Staphylococcus aureus* skin infections in mice. *ChemMedChem* **9**, 710-713 (2014).
- Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J.** 16S Ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* **173**, 697-703 (1991).
- Weissman K. J. & Müller, R.** Myxobacterial secondary metabolites: Bioactivities and modes-of-action. *Nat. Prod. Rep.* **27**, 1276-1295 (2010).
- Wells, J. S., Hunter, J. C., Astle, G. L., Sherwood, J. C., Ricca, C. M., Trejo, W. H., Bonner, D. P. & Sykes, R. B.** Distribution of β -lactam and β -lactone producing bacteria in nature. *J. Antibiot.* **35**, 814-821 (1982).
- Wells, J. S., Trejo, W. H., Principe, P. A. & Sykes, R. B.** Obaf fluorin, a novel β -lactone produced by *Pseudomonas fluorescens*. Taxonomy, fermentation and biological properties. *J. Antibiot.* **37**, 802-803 (1984).

- Wiegand, I., Hilpert, K. & Hancock, R. E.** Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **32**, 163-175 (2008).
- Williams, K. P.** Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.* **30**, 866-875 (2002).
- Williamson, N. R., Fineran, P. C., Leeper, F. J & Salmonds, G. P. C.** The biosynthesis and regulation of bacterial prodiginines. *Nat. Rev. Microbiol.* **4**, 887-899 (2006).
- Wilson, M. C. & Piel, J.** Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem. Biol.* **20**, 636-647 (2013).
- Wilson, M. C., Mori, T., Rückert, C., Uria, A. R., Helf, M. J., Takada, K., Gernert, C., Steffens, U. A. E., Heycke, N. & others.** An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58-26 (2014).
- Winkler, R. & Hertweck, C.** Biosynthesis of nitro compounds. *ChemBioChem* **8**, 973-977 (2007).
- Winstanley, C., Langille, M. G. I., Fothergill, J. L., Kukavica-Ibrulj, I., Paradis-Bleau, C., Sanschagrin, F., Thomson, N. R., Winsor, G. L., Quail, M. A. & others.** Newly introduced genomic prophage islands are critical determinants of *in vivo* competitiveness in the Liverpool epidemic strain of *Pseudomonas aeruginosa*. *Genome Res.* **19**, 12-23 (2009).
- Wissing, F.** Cyanide formation from oxidation of glycine by a *Pseudomonas* species. *J. Bacteriol.* **117**, 1289-1294 (1974).
- Wolf, F., Bauer, J. S., Bendel, T. M., Kulik, A., Kalinowski, J., Gross, H. & Kaysser, L.** Biosynthesis of the β -lactone proteasome inhibitors belactosin and cystargolide. *Angew. Chem. Int. Ed.* **56**, 6665-6668 (2017).
- Wolpert, M., Gust, B., Kammerer, B. & Heide, L.** Effects of deletions of *MbtH*-like genes on chlorobiocin biosynthesis in *Streptomyces coelicolor*. *Microbiology* **153**, 1413-1423 (2007).
- Wu, K., Chung, L., Revill, W. P., Katz, L. & Reeves, C. D.** The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascofeticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* **251**, 81-90 (2000).
- Wu, J., Cooper, S. M., Cox, R. J., Crosby, J., Crump, M. P., Hothersall, J., Simpson, T. J., Thomas, C. M. & Willis, C. L.** Mupirocin H, a novel metabolite resulting from mutation of the HMG-CoA synthase analogue, *mupH* in *Pseudomonas fluorescens*. *Chem. Commun.* 2040-2042 (2007).
- Wuest, M. W., Sattely, E. S. & Walsh, C. T.** Three siderophores from one bacterial enzymatic assembly line. *J. Am. Chem. Soc.* **131**, 5056-5057 (2009).
- Wyatt, M. A., Ahilan, Y., Argyropoulos, P., Boddy, C. N., Magarvey, N. A. & Harrison, P. H.** Biosynthesis of ebelactone A: isotopic tracer, advanced precursor and genetic studies reveal a thioesterase-independent cyclization to give a polyketide β -lactone. *J. Antibiot.* **66**, 421-430 (2013).

- Yakasai, A. A., Davison, J., Wasil, Z., Halo, L. M., Butts, C. P., Lazarus, C. M., Bailey, A. M., Simpson, T. J. & Cox, R. J.** Nongenetic reprogramming of a fungal highly reducing polyketide synthase. *J. Am. Chem. Soc.* **133**, 10990-10998 (2011).
- Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E. & Ikeda, H.** Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. USA* **112**, 857-862 (2015).
- Yang, H. W. & Romo, D.** Methods for the synthesis of optically active β -lactones (2-oxetanones). *Tetrahedron* **55**, 6403-6436 (1999).
- Yin, P. -J., Wang, J. -S., Wang, P. -R. & Kong, L. -Y.** Sesquiterpenes and lignans from the fruits of *Illicium simonsii* and their cytotoxicities. *Chin. J. Nat. Med.* **10**, 383-387 (2012).
- Yorgey, P., Lee, J., Kördel, J., Vivas, E., Warner, P., Jebaratnam, D. & Kolter, R.** Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. *Proc. Natl. Acad. Sci. USA* **91**, 4519-4523 (1994).
- Yoshinari, K., Aoki, M., Ohtsuka, T., Nakayama, N., Itezono, Y., Mutoh, M., Watanabe, J. & Yokose, K.** Panlicins, novel pancreatic lipase inhibitors. II. Structural elucidation. *J. Antibiot.* **47**, 1376-1384 (1994).
- Yu, T. W., Bai, L., Clade, D., Hoffman, S., Toezler, S., Trinh, K. Q., Xu, J., Moss, S. J., Leistner, E. and Floss, H. G.** The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*. *Proc. Natl. Acad. Sci. USA* **99**, 7968-7973 (2002).
- Zachow, C., Jahanshah, G., de Bruijn, I., Song, C., Ianni, F., Pataj, Z., Gerhardt, H., Pianet, I., Lämmerhofer, M., Berg, G., Gross, H. & Raaijmakers, J. M.** The novel lipopeptide poeamide of the endophyte *Pseudomonas poae* RE*1-1-14 is involved in pathogen suppression and root colonization. *Mol. Plant Microbe Interact.* **28**, 800-810 (2015).
- Zarins-Tutt, J. S., Barberi, T. T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D. J. & Goss, R. J. M.** Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial *cryptic* or *silent* natural products. *Nat. Prod. Rep.* **33**, 54-72 (2015).
- Zelyas, N. J., Cai, H., Kwong, T. & Jensen, S. E.** Alanylclavam biosynthetic genes are clustered together with one group of clavulanic acid biosynthetic genes in *Streptomyces clavuligerus*. *J. Bacteriol.* **190**, 7957-7965 (2008).
- Zerikly, M. & Challis, G. L.** Strategies for the discovery of new natural products by genome mining. *ChemBioChem* **10**, 625-633 (2009).
- Zha, W., Rubin-Patel, S. B. & Zhao, H.** Characterization of the substrate specificity of PhID, a type III polyketide synthase from *Pseudomonas fluorescens*. *J. Biol. Chem.* **281**, 32036-32047 (2006).
- Zhao, L., Sherman, D. H. & Liu, H. -W.** Biosynthesis of desosamine: construction of a new methymycin/neomethymycin analogue by deletion of a desosamine biosynthetic gene. *J. Am. Chem. Soc.* **120**, 10256-10257 (1998).

- Zhang, Y., Rowley, K. B. & Patil, S. S.** Genetic organization of a cluster of genes involved in the production of phaseolotoxin, a toxin produced by *Pseudomonas syringae* pv. *phaseolicola*. *J. Bacteriol.* **175**, 6451-6458 (1993).
- Zhang, W., Ostash, B. & Walsh, C. T.** Identification of the biosynthetic gene cluster for the pacidamycin group of peptidyl nucleoside antibiotics. *Proc. Natl. Acad. Sci. USA* **107**, 16828-26833 (2010).
- Zhang, W., Heemstra, J. R., Walsh, C. T. & Imker, H. J.** Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. *Biochemistry* **49**, 9946-9947 (2010).
- Zhang, G., Zhang, H., Li, S., Xiao, J., Zhang, G., Zhu, Y., Niu, S., Ju, J. & Zhang, C.** Characterization of the amicetin biosynthesis gene cluster from *Streptomyces vinaceusdrappus* NRRL 2363 implicates two alternative strategies for amide bond formation. *Appl. Environ. Microbiol.* **78**, 2393-2401 (2012).
- Zhang, C., Kong, L., Liu, Q., Lei, X., Zhu, T., Yin, J., Lin, B., Deng, Z. & You, D.** *In vitro* characterization of echinomycin biosynthesis: formation and hydroxylation of L-tryptophanyl-S-enzyme and oxidation of (2S,3S) β -hydroxytryptophan. *PLoS ONE* **8**, e56772 (2013).
- Zhang, M. M., Wang, Y., Ang, E. L. and Zhao, H.** Engineering microbial hosts for production of bacterial natural products. *Nat. Prod. Rep.* **33**, 963-987 (2016).
- Zhao, C., Ju, J., Christenson, S. D., Smith, W. C., Song, D., Zhou, X., Shen, B. & Deng, Z.** Utilization of the methoxymalonyl-acyl carrier protein biosynthesis locus for cloning the oxazolomycin biosynthetic gene cluster from *Streptomyces albus* JA3453. *J. Bacteriol.* **188**, 4142-4147 (2006).
- Zhao, C., Coughlin, J. M., Ju, J., Zhu, D., Zhu, D., Wendt-Pienkowski, E., Zhou, X., Wang, Z., Shen, B. & Deng, Z.** Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an "Acyltransferase-less" type I polyketide synthase that incorporates two distinct extender units. *J. Biol. Chem.* **285**, 20097-20108 (2010).
- Ziemert, N., Alanjary, M. & Weber, T.** The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.* **33**, 988-1005 (2016).
- Zou, J., Pan, L., Li, Q., Zhao, J., Pu, J., Yao, P., Gong, N., Lu, Y., Kondratyuk, T. P. & others.** Rubesanolides A and B: Diterpenoids from *Isodon rubescens*. *Org. Lett.* **13**, 1406-1409 (2011).
- Zou, J., Pan, L., Li, Q., Pu, J., Yao, P. Zhu, M., Banas, J. A., Zhang, H. & Sun, H.** Rubesanolides C-E: abietane diterpenoids isolated from *Isodon rubescens* and evaluation of their anti-biofilm activity. *Org. Biomol. Chem.* **10**, 5039-5044 (2012).

Appendix 1:
Supplementary Figures and
Tables

APPENDIX 1: Supplementary Figures and Tables

PaGlyA1	1	MFSKHDQIRGYDDELLAAMDAAEEARQEEHIELIASENYTSKRVMQAQGSGLTNKYAEGYP
PfGlyA	1	MFSRDLTIAKYDADLFAAMEQEAQRQEEHIELIASENYTSQAVMEQAQGSVLTNKYAEGYP
1EQB_B	1	MLKREMNADYDAELNQAAMEQEKVQEEHIELIASENYTSQPRVMAQAGSGLTNKYAEGYP
2VMV_A	1	----MKYIPQQDPQVFAATEEQERKRQCHAKIELIASENYTSRAVMEQAQGSVLTNKYAEGYP
ObaG	1	MSNVKQQTAAQIVDWLSSTLGGKDHQYRELSLSLTANENYPSALVRLTSGSTAGAFYHCSFP
PaGlyA1	61	----GKRYYGGCEHVDKVERLAIDRAEQLEFGADYANVQPHSGSSANAAVYLALLNAGDT
PfGlyA	61	----GKRYYGGCEYVDNVEQLAIDRAKELFGADYANVQPHAGSQANSAAVYLALLQGGDT
1EQB_B	61	----GKRYYGGCEYVDNVEQLAIDRAKELFGADYANVQPHSGSQANFAVYTTALLEPGDT
2VMV_A	57	----GKRYYGGCEYVDNVEELARETRAKQLFGAEHANVQPHSGAQANMAVYFTVLEHGDT
ObaG	61	FEVPAQGEWHFPEPGHMAADQVRDLKRTLLIQAQAFDWRENGSGSTAEQAAMLAACKPGE
PaGlyA1	116	ILGMSLAHGGHLTHGARVSSSGKLYNAVQYGLDTATGLIDYDEVERLAVEH-KPKMIVAG
PfGlyA	116	ILGMSLAHGGHLTHGASVSSSGKLYNAVQYGLD-ANGLIDYDEVERLAVEH-KPKMIVAG
1EQB_B	116	VLGMNLAHGGHLTHGSPVNFSGKLYNIVPYGID-ATGHIDYADLEKQAEH-KPKMIIGG
2VMV_A	112	VLGMNLSHGGHLTHGSPVNFSGVQYNFVAYGVDPETHVIDYDDVREKARLH-RPKLIVAA
ObaG	121	FVHFHRDGGHFALESLSACKMG--IEIFHLFVNPTSLLLIDVAKLDEMRNPHIRIVILD
PaGlyA1	175	FSAYSKTIDFPRFRAIADKVGALLFVDMAHVAGLVAAGLYPNPIPF-ADVVTTHHTKTLR
PfGlyA	174	FSAYSQIIDFPRFRAIADKVGAYLFVDMAHVAGLVAAGVYPNPVPI-ADVVTTHHTKTLR
1EQB_B	174	FSAYSGVVDIAKREIADSI GAYLFVDMAHVAGLVAAGVYPNPVPH-AHVVTTHHTKTLA
2VMV_A	171	ASAYPRIIDFAKFRITADEVGAYLMVDMAHVAGLVAAGLHPNPVPI-AHFVTTHHTKTLR
ObaG	179	QSFKLRWQPLAEIRSMLPDS-CTLTIDYDMSHDGGLIMGVVDSPLSCGADIVHGNTHTKTP
PaGlyA1	234	GPRGGLILAR-ANAELEKKLNSAVFPGAQGGPLMHVIAAKAVCFKEALEPFGFKDYQAQVI
PfGlyA	233	GPRGGLILAR-ANAELEKKLNSAVFPGAQGGPLEHVIAAKAICFKEALQEPFVKTYQQQVV
1EQB_B	233	GPRGGLILAKGGSEELYKKLNSAVFPGAQGGPLMHVIAKAVALKAELEPFGFKTYQQQVA
2VMV_A	230	GPRGGVILCQ---EQFAKQIDKAI FPGI QGGPLMHVIAAKAVAFGEALQDFKAYAKRVV
ObaG	238	GPQKVIIGFKSAQHPMLVDTSLWVCPHLQSNCHAEQPPMWVAFKEMELFG-RDYAAQIV
PaGlyA1	293	RNAKAMAEVFI GRGYD VVSSG---TDNHLMLTSLVRQGLTGKEADAALGRVGITVNKNVA
PfGlyA	292	KNAQTMASVFI ERGFD VVSSG---TDNHLMLTSLIKODISGKDADAALGKAFITVNKNSV
1EQB_B	293	KNAKAMVEVFL ERGRV VVSSG---TDNHLMLVDLVDKNTLGKEADAALGRANITVNKNSV
2VMV_A	287	DNAKRMASALQNEGFTV VSSG---TDNHLMLVDLRFQGLTGKTAEKVLEDEVGITVNKNIT
ObaG	297	SNAKTLARHLHELGLDVTGESFSGFTQTHQVHFAVGDLCALDLCVNSLHAGGIRSTNIEI
PaGlyA1	350	PNDPQSPFVTSGIRIGTPAITTRGLOEAQSRRELAWICDILDLHIGDADVEAKVATQVAGL
PfGlyA	349	PNDPQSPFVTSGIRIGTPAITTRGFKEAECKELAGWICDILADLNNEAVIDAVREKVKAI
1EQB_B	350	PNDPQSPFVTSGIRIGTPAITTRGFKEAEAKELAGWICDVLDSINDEAVIERIKGKVLDI
2VMV_A	344	PYDPESEPCVTSGIRIGTAAVTRGFLEEMDEIAAIIGLVILKNVGSEQALBEARQVAAL
ObaG	357	PGKPG----VHGIRLGVQAMTRRGMKELFEVVARFIADLYFKKTEPAKVAQQIKEFLQA
PaGlyA1	410	CADFVYR-----
PfGlyA	409	CKKLPVYGA-----
1EQB_B	410	CARYPVYA-----
2VMV_A	404	TD-----
ObaG	413	FPLAFLAYSFDNYLDEELLA AVYQGAQR

Supplementary Figure 1. Sequence alignment of authentic and putative SHMT amino acid sequences. ObaG from the *oba* BGC is aligned (ClustalX2) with predicted SHMTs from *P. aeruginosa* PA01 (PaGlyA1, NP_254102.1) and *P. fluorescens* SBW25 (PfGlyA, WP_015885982.1) whole-genome sequences, and validated SHMTs with solved structures from *E. coli* (EcGlyA, 1EQB_B) and *B. stearothermophilus* (BsGlyA, 2VMV_A). ▼ indicates the conserved lysine residue required for internal aldimine formation with PLP.

```

LipK      1  -----MTVGAGGKTSADADPLMVRATADADRRRAAFAINLVPSENRISPLASLPLA
FTase     1  MPSSVNRTSRTEPAGHHREFPLSLAATDELVAEEEAEDAFLVHLTANETVLSFRARAVLA
ObaG      1  -----MSNVKQQTAAQIVDWLSSTLGKDHQYREDSLSLTANENYPSALVRLTSG

LipK      52  SDFYNRYFFN-----TDCDPLFWFRGGEDIAHIEALCAAALRRMASARYCNVRPISG
FTase     61  SPLTSRYLLEHLDMRGPSPARLGNLLLRGLDRIGTIEESATEVCCRRLFARYAEFRCLSG
ObaG      49  STAGAFVHCS-----FPEVVPAGEWHFPEPGHMNATAQVRDILGKTLIGAQAFDWRPNGG

LipK      105  MSAMILTVAALSPPGSTVVSVDQNSGGHATPALLGRIGRRSRLLNCK-DGE-VDESELA
FTase     121  LHAMQTFHAALSRRPGDTVMRVATKGGHLELILCRSFGRRSCTYVFD-DFMTIDLERTR
ObaG      104  STAEQALMLAACKPGEGFVHFAHRDGGHFALES LAQKMGIEIFHLPVNPTSLIDVAKLD

LipK      163  EVLAPG-DVALVYVDVQNCVRVPDFRRSDVIREVSPGTRLYVDASHYLGVLGGLLANP
FTase     180  EVVEKE-FPSLLFVDAMNYLFFFPFAELKATAGDVP----LVFDASHTLGLTAGGRFQDP
ObaG      164  EMVRRNPHIRIVIIDQSFKLRWQPLAELRSVLPDSCT---LTYDMSHDGLLVGGVFDSP

          ▼

LipK      222  LDCGDAFGGSGTHKSFPGPHKGVIFTN--AEDYDESLR-SAQFDLVSSHHFAETLALSIA
FTase     235  LREGADLIQANTHKTFEGPQKGIILGN--DFSLMEELGYTLSTGMVSSQHTASTVALLIA
ObaG      221  LSCGADIVHGNTHKTPGPQKGYIGFKSAQHPFLVDTSLWVCPHLOSNCHEAQLPPMWA

LipK      279  ALEVEDRMGDYARATNDNARRLAGALADAGFRVYGDSATGNTDTHQVWVEIDGVAAAYAL
FTase     293  LHEMWDGRYAAQVIDNARRLAGALRDRGVPVVAE-ERGFTANHMFEVDTRPLGSGPAV
ObaG      281  FKEMELFGRDYAAQIVSNAKTLARHLHELGLDVTGE-SFGFTQTHQVHFAVGDILQKALDL

LipK      339  SNR-LAEGGIRVNLQSSMPGMSGVHLRLGSNEVTFEGAGPQATEELAGALVTARER----
FTase     352  IQR-LVRAGVMSANRAVAFNHLID--TIRFGVQETRRGYDHDDLDEAADLVAAVLLERQEP
ObaG      340  CVNSLHAGGIRSTNIEIPKPGVHGIIRLVQAMTRRGMKEKDFEVVARFLADLYFKKTEP

LipK      394  -ALGPRTVHEIRCRFGAPFYTDPEKLVVACI-----
FTase     409  ERIRPRVAELVGRRTVRVTGDPASAAGPPARERYAPPTAPAGHPARPRWIGVRLTPLPE
ObaG      400  AKVAQQIKKFLQAFPLAFLAYSFDNYLDEELTAAVYQGAQR-----

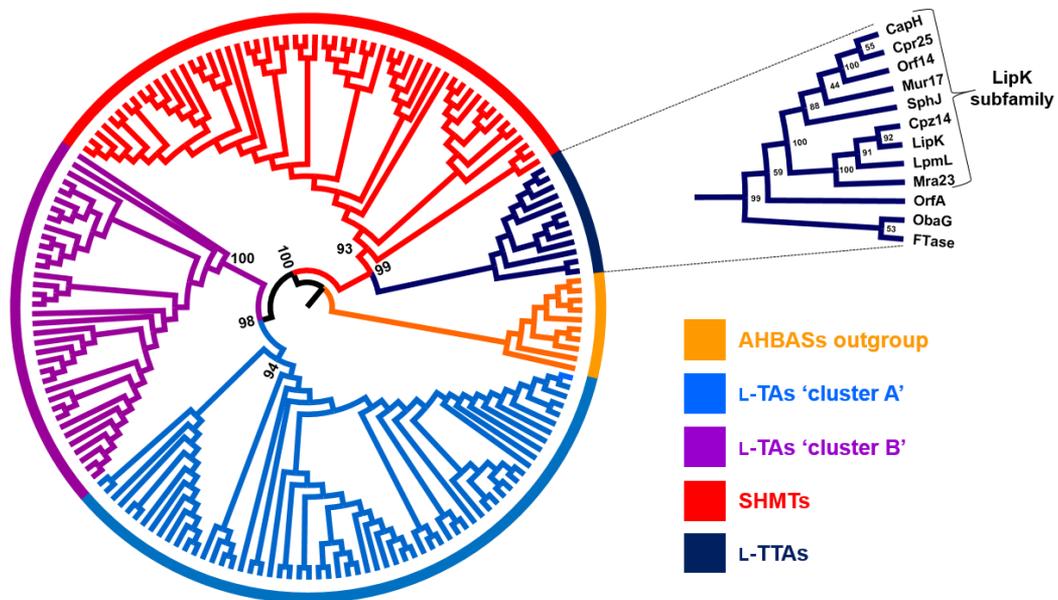
LipK      -----
FTase     469  PVTEAECAGAQLRGLAGAFPHQIDSSGNVSFTSTDGRLFVTGSGTYIKDLAPGDFVELT
ObaG      -----

LipK      -----
FTase     529  GAEGWTLHCRGDGPPSAEAYLHLLLRERVGARYVVHNCIPGRALETSGALVIPKEYGS
ObaG      -----

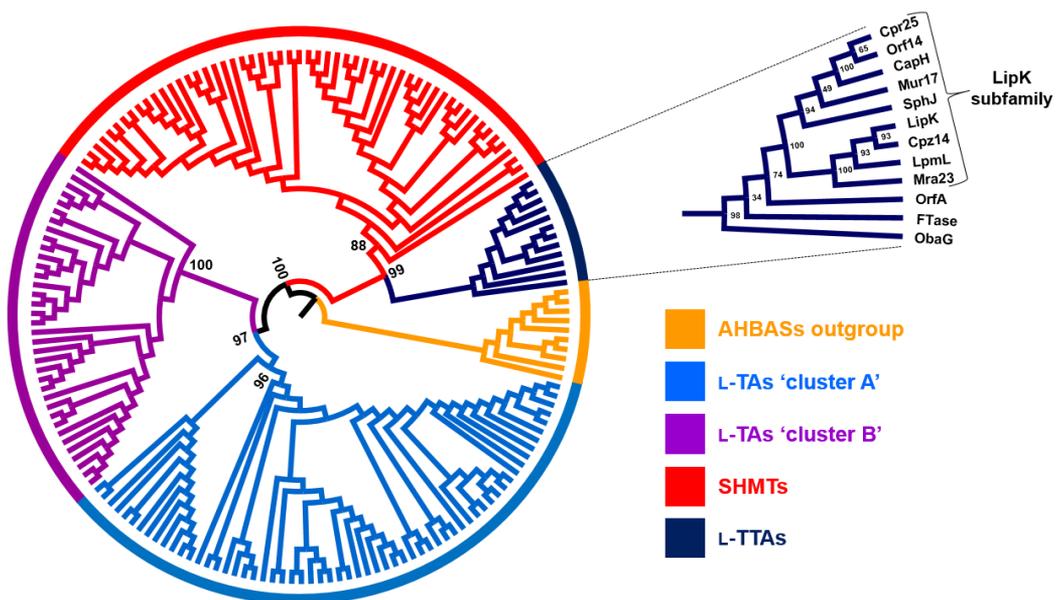
LipK      -----
FTase     589  VALAEAVADACQDSQVMYVRRHGLVFWAHSYDECLALIEDVRRITG
ObaG      -----

```

Supplementary Figure 2. Sequence alignment of biochemically characterised L-TTAs. ObaG from the *oba* BGC is aligned (ClustalX2) with 4-FTase from *Streptomyces cattleya* (WP_014151017.1)⁷ and LipK from *S. sp.* SANK 60405 (BAJ05887.1). FTase comprises an N-terminal SHMT-like domain and a C-terminal phosphate binding domain similar to those of bacterial aldolases and empimerases. ObaG shares 29% identity (88% query cover) to the N-terminal domain of FTase, and 33% identity (67% query cover) with LipK. ▼ indicates the conserved lysine residue required for internal aldimine formation with PLP.



Supplementary Figure 3. Maximum likelihood tree of SHMTs, L-TAs and L-TTAs version 2. A set of 3-amino-5-hydroxybenzoate synthase (AHBAS) amino acid sequences serve as the outgroup. RAxML likelihood values at the root nodes for SHMT and L-TA cluster clades are annotated. The clade comprising the L-TTA ObaG characterised in this work has been expanded and annotated with RAxML likelihood values and the identity of amino acid sequences represented. Initial 'conservative' trimming performed following alignment in ClustalX2.



Supplementary Figure 4. Maximum likelihood tree of SHMTs, L-TAs and L-TTAs version 3. A set of 3-amino-5-hydroxybenzoate synthase (AHBAS) amino acid sequences serve as the outgroup. RAxML likelihood values at the root nodes for SHMT and L-TA cluster clades are annotated. The clade comprising the L-TTA ObaG characterised in this work has been expanded and annotated with RAxML likelihood values and the identity of amino acid sequences represented. Initial trimming performed following alignment in MUSCLE.

1QF6 1 --MPVITLPGDSQRHVDHAVSPMDVALDIGPGLAKACIAGRVNGELVDACDLIENDAQS
 1NYQ 1 MEQINIQFPDGNKAFDKGTTTEDIAQSISPLGRKKAAGKFNGLVDLTKPLETDGSGIE
 4P3N 1

 39502'S' 1 --MPVITLPGDSQRSFDHPVSVAEVAASIGAGLAKATVAGKVDGQLVDASDLITSDASIQ
 obaO 1 --MVTIALPDGSRRDPEALITVQQLAOSIGAGLAAATCGKVDGTLVDASYLLETDAIVE
 Cshin'S' 1 --MPVITLPGDSQRQFDAPVTVAGVAASIGAGLAAATAGKVDGKLVDTSYLIDRDVQLA
 Cshin'R' 1 --MITISLPGDSKREFAEPI SVHELACAI GPGLCAAATAGKVDGKLVDTAHLRHDATVE
 Bdiff'S' 1 --MVSIRLPDGSVROYEH PVTVAEVAASIGPGLAKAALGGKIDGELVDTSTVIDRDASTA
 Bdiff'R' 1 --MISTIALPDGSRRAVDHPVTVAALAADIGPGLAKAALAGKIDGKLVLDYLLIDIDAFAE
 Pf'S' 1 --MPVITLPGDSQRSFDHPVSVAEVAASIGAGLAKATVAGKVDGQLVDASDLITSDASIQ
 Pf'R' 1 --MVTIALPDGSRRDPEALITVQQLAOSIGAGLAAATCGKVDGTLVDASYLLETDAIVE

1QF6 59 IITAKDEEGLEIIRHSCAHL LGHALKQLVPHTKVAIGPVIDNGFYFDVLDLRTLTQEDME
 1NYQ 61 IVTPGSEEALEVIIRHSTAHLMAHAIKRLYGNVKFGVGPVIEGGFYFDIDIQNISSDDFE
 4P3N 1

 39502'S' 59 IITPKDQEGLEIIRHSCAHLIGHAVKQOLYPTAKVIGPVIDEGFYFDIAYERPFTPDDLA
 obaO 59 IVTKSPQALEIIRHSTAHLMAQAVQRLYPGTQVTIGPVIDNGFYFDVAPRPFTMDDL P
 Cshin'S' 59 IVTEFDADGLDVIIRHSTAHLAYAVKELSSAQVTIGPVIDEGFYFDVSYERPFTPEDLD
 Cshin'R' 59 IVTDHHPDALEVIRHSTAHLAQAVQRLYPGTQVTIGPVIDNGFYFDVAGERPFTVEDLP
 Bdiff'S' 59 IVTDKADGLDIIRHSTAHLAYAVKDLYPDAQVTIGPVIDNGFYFDVSYNRPFTPEDLE
 Bdiff'R' 59 IVTEKHPDALSIIRHSCAHLAQAVQRLYPAAQFSIGPVIDNGFYFDVTSISPPLESDDL P
 Pf'S' 59 IITPKDQEGLEIIRHSCAHLIGHAVKQOLYPTAKVIGPVIDEGFYFDIAYERPFTPDDLA
 Pf'R' 59 IVTKSPQALEIIRHSTAHLMAQAVQRLYPGTQVTIGPVIDNGFYFDVAPRPFTMDDL P

1QF6 119 ALEKRMHELAEKNDYVTKKVSWEARETFANRCESYKVSILDENIAHDKPGLYFHHEM
 1NYQ 121 QIEKTMKQIVNENKIERKVVSRDEAKELFSN--DEYKLELIDA-IPEDENVTLYSQGEF
 4P3N 1

 39502'S' 119 AIEQRMHALIEKDYDVTKKVTPRAEVIVDVFARGEDYKLRIVED-IPDEQAMGLYHHEM
 obaO 119 LIEABMTRIVKEQLPVRTQIPRDEALAFEEQLGESYKTIIDA-IPAGETLSLYIQGEF
 Cshin'S' 119 AIEKKMFELSKKDI PVERYELSRDAIAYFKGIGEAYKAEIIES-IPQNEVLSLYREGGF
 Cshin'R' 119 AIEAEMARITAKEALPVTRSEKTRQAAQFEGLGEHYKVEILRD-IADDQPLSLYIQGEF
 Bdiff'S' 119 KIEKRMQELAKKDEPVTRRVSRDEAAGYFRSLGEKYKAEIIES-IPQSEIEIKLYSHGGF
 Bdiff'R' 119 RIEABMRAIVAEAIPVSRVLSRDAIREFSDRGQTYKAEIVAS-IPPEHQLTIIYIQGEF
 Pf'S' 119 AIEQRMHALIEKDYDVTKKVTPRAEVIVDVFARGEDYKLRIVED-IPDEQAMGLYHHEM
 Pf'R' 119 LIEABMTRIVKEQLPVRTQIPRDEALAFEEQLGESYKTIIDA-IPAGETLSLYIQGEF

1QF6 179 VDLCRGPHVPMRFC--HFKLMKKTAGAYWRGDSNNKMLQRIYGTAWADKKALNAYLORLEE
 1NYQ 178 TDLCRGVHVPSTAKLKEFKLISTAGAYWRGDSNNKMLQRIYGTAFDFKKELKAHLQMLEE
 4P3N 1 -----MGSS-----HHHHH

 39502'S' 178 VDLCRGPHVPNTRFLMSFKLTKLSGAYWRGDAKNEQLQRIYGTAWADKKOLAAYIORLEE
 obaO 178 TDLCRGPHVPNTAKLGAFLMKVAGAYWRGDSNNIMLSRIYGTAWGNEKELKAYLNQLOE
 Cshin'S' 178 TDLCRGPHVPSTKLVKFKLMKVAGAYWRGDSRNEMLTRVYGTAWAKKELDAYLHRLEE
 Cshin'R' 178 TDLCRGPHVPNTKLVKFKLMKVAGAYWRGNSDNAMLSRIYGTAWLNDKDLKAYLLOLEE
 Bdiff'S' 178 TDLCRGPHVPSTKLVKFKLMKVAGAYWRGDSKNEQLQRIYGTAWTKKELQDQYLHMLEE
 Bdiff'R' 178 TDLCRGPHVPNTRALFAFKLMKKTAGAYWRGDSNNEMLCRVYGTAWLNDALIQAYLHQIAE
 Pf'S' 178 VDLCRGPHVPNTRFLMSFKLTKLSGAYWRGDAKNEQLQRIYGTAWADKKOLAAYIORLEE
 Pf'R' 178 TDLCRGPHVPNTAKLGAFLMKVAGAYWRGDSNNIMLSRIYGTAWGNEKELKAYLNQLOE

1QF6 239 AAKRDHRKIGKQLDLNHH--QEEAPGMVFWHNDGWTIERELEVFRSKLKEYOYQEVKGF
 1NYQ 238 RKE RDHRKIGKELDLFTNSQLVCAGLPLWLPNGATIRRETERYIVDKEVSMGYDHYVTPV
 4P3N 10 HSSGDHRKIGRDQDLNFF--HELSPGSCFLLPKGAYLYNALIEFIRSEYRKRGFQEVVTPN

 39502'S' 238 AEKRDHRKIGKRLNLFHL--QEEAPGMVFWHPNGWILVQVLEQYMRVQRENGYLEIKTPQ
 obaO 238 AEKRDHRKIAKQFDLFHQ--QEEAPGMVFWHPKGWSLWQVVEQYMRVYREGGYREVKSPQ
 Cshin'S' 238 AEKRDHRKIGKALDLFHM--QEEAPGMVFWHPKGWSLWQVVEQYTRRLAKAGYQEVVTPM
 Cshin'R' 238 AEKRDHRKIAKLLDLFHQ--QEEAPGMVFWHYKGWALWQAVEQYMRVYRDSGYREVKAPO
 Bdiff'S' 238 AEKRDHRKLGKQLDLFHM--QEEAPGMVFWHPKGWALWQVEQYMRVRSVAGYLEIKTPM
 Bdiff'R' 238 AE RDHRKIGKQLDLFHI--QEEAPGMVFWHPKGWSLWQVVEQYMRVYVVECGYQEVKAPQ
 Pf'S' 238 AEKRDHRKIGKRLNLFHL--QEEAPGMVFWHPNGWILVQVLEQYMRVQRENGYLEIKTPQ
 Pf'R' 238 AEKRDHRKIAKQFDLFHQ--QEEAPGMVFWHPKGWSLWQVVEQYMRVYREGGYREVKSPQ

▼

1QF6	298	MMDRVLWEKIGHWQNYKQAMFVTS--SENREYCIKPMNCPGHVQVFNQGLKSYRDLPLRM
1NYQ	298	LANVDLYKTSQHWQHYQEDMFPPMQLDETESMVLKPMNCPHHVMTYANKPHSYRELPIRI
4P3N	69	LFNRLWMTSGHWQHYSENMFVFE--VEKELFALKPMNCPGHCLMFDHRFRSRELPIRL
39502'S'	297	VDRSLWEKSGHWANYADNMFTTQ--SENRDYAIKPMNCPCHVQVFNQGLKSYRELPIRL
obaO	297	VLDSLWKKSGHWQNYKENMFVTE--SENROYALKPMNCPGHIQIFKHGLRSHRELPIRY
Cshin'S'	297	MMDRSLWEKSGHWQNYQENMFITE--SEKRTYAIKPMNCPGHIQIFSSDLRSYRDLPLRL
Cshin'R'	297	VLDVSLWQKSGHWQNYQENMFLTE--SEKROYALKPMNCPGHIQIFKQGLRSYRELPIRY
Bdiff'S'	297	IMDRSLWEASQHWQNYRENMFVTE--SEKRDYAIKPMNCPGHVQVFKHGLRSYRDLPLRY
Bdiff'R'	297	VVDVSLWKRSGHWQNYKENMFVTE--SEKREYALKPMNCPGHIQIFKHGLRSYRDLPLRY
Pf'S'	297	VDRSLWEKSGHWANYADNMFTTQ--SENRDYAIKPMNCPCHVQVFNQGLKSYRELPIRL
Pf'R'	297	VLDSLWKKSGHWQNYKENMFVTE--SENROYALKPMNCPGHIQIFKHGLRSHRELPIRY

▼▼ ▼ ▼▼▼

1QF6	356	AEEFGSCHRNEPSGSLHGIMRVRGFTQDDAHIFCTEEQIRDEVNCGCIRIVYDMYSTFGFEK
1NYQ	358	AELGTMHRYEASGALSGLQVRVGMTLNDSHIFVRPQIQKEEFKRVVNIIDVYKDFGFEED
4P3N	127	AEEFVLRHRELSGALTGLTRVRRFQDDAHIFCAMEQIEDEIKGCCLDFLRTVYVSVFGFSF
39502'S'	355	AEEFGCHRNEPSGALHGIMRVRGFTQDDAHIFCTEEQIAEESAFAIKITMDVYRDFGFTD
obaO	355	GEFGCHRNEPSGALHGIMRVRAFTQDDGHIFCTEEQIAAEIKAFHYQAKVYADFGFTD
Cshin'S'	355	AEEFGCHRNEPSGALHGIMRVRGFTQDDAHIFCTEQLIDEARIFHAIAMSVYDDFGFEG
Cshin'R'	355	GEFGCHRNEPSGALHGIMRVRAFTQDDGHVFCCTEEQIADEVQAFHRQAKVYADFGFDN
Bdiff'S'	355	AEEFGSCHRNEASGALHGIMRVRGFTQDDAHIFCTEEQIFIAESIAFNITAMSVYKDFGFEH
Bdiff'R'	355	GEFGCHRNEASGALHGIMRVRAFTQDDGHIFCTEQLIDEVAAFHRQAKVYRDFGFGD
Pf'S'	355	AEEFGCHRNEPSGALHGIMRVRGFTQDDAHIFCTEEQIAEESAFAIKITMDVYRDFGFTD
Pf'R'	355	GEFGCHRNEPSGALHGIMRVRAFTQDDGHIFCTEEQIAAEIKAFHYQAKVYADFGFTD

▼

1QF6	416	--IVVKLSTR--PEK--RIGSDEIWDRAEADLAVALEENNIPEFYQLGEGAFYGPKIEET
1NYQ	418	--YSEFLSYRDPKKEYFDDDDWNAENMLKEAADELGLSYEEAIGEAIFYGPKLDVQ
4P3N	187	---KINLSTR--PEK--FLGDIEIWDQAEKQLENSINEFGEKWELENSGEGAFYGPKI IQ
39502'S'	415	--VENKLSTR--PEK--RVGSDEIWDRAEAALAAALDSAGLAYDLQPGEGAFYGPKIEES
obaO	415	--IANKLALR--PEPGKRLGSDEIWDKAENLLREALSECDVWEWELPGEAFYSPKIEYH
Cshin'S'	415	--IANKLALR--PEK--RAGSDEIWDRAEHLREALRACGVEWEELPGEAFYGPKIEYH
Cshin'R'	415	--IANKLALR--PEAGKRLGSDEIWDRAEALLRSALSACGVTWEELPGEAFYSPKIEYH
Bdiff'S'	415	--IDIKLSLR--PEQ--RAGIDETIWDRAEQGLRDALTACGLSWEELPGEAFYGPKIEYH
Bdiff'R'	415	DSIANKLALR--PEL--RIGSDEIWDRAENTLRDALRKCGVEWEELPGEAFYSPKIEYH
Pf'S'	415	--VENKLSTR--PEK--RVGSDEIWDRAEAALAAALDSAGLAYDLQPGEGAFYGPKIEES
Pf'R'	415	--IANKLALR--PEPGKRLGSDEIWDKAENLLREALSECDVWEWELPGEAFYSPKIEYH

▼▼▼ ▼▼ ▼

1QF6	470	LYDCIDRAWQCCTQLDFSLPSRLSASVYGEDNE--RIVPVMITHRAIGSMERFIGILIEE
1NYQ	476	NKTANGKEETLSTAQLDELLPERFDLTYIGQDGE--HHPVVIHRGVSMERFVAFLEIEE
4P3N	240	IKDAIGRYHQCATQLDFQLPIRFNLTYVSHDGGDKKRPVIVHRAIGSVERMIAILIEE
39502'S'	469	LKDCIGRVWQCCTQLDFNLPIRLGAEYVVEDNS--RKHVPMMLHRAIGSFERFVILIEH
obaO	471	LRDAIGREWQVGTQDYLHMPDRLGAEYVDEHSQ--RRKPVMLHRAIVGSMERFLGILIEH
Cshin'S'	469	IKDAIGRSWQCCTQLDFVLPERFGEYVAEDNS--RQSPVMLHRAIVGSMERFLGILIEH
Cshin'R'	471	LKDAIGREWQVGTQDYLMPERLGAEYVDEHSQ--RRSPVMLHRAIVGSMERFIGILIEH
Bdiff'S'	469	IKDAIGRSWQCCTQLDMVLPERLGAEYVAEDNS--RRFPVMLHRAIVGSMERFLGILIEH
Bdiff'R'	471	LKDAIGREWQVGTQADYLMPEKLGAEYVDDRSE--RRTPVMLHRAIVGSMERFIGILIEH
Pf'S'	469	LKDCIGRVWQCCTQLDFNLPIRLGAEYVVEDNS--RKHVPMMLHRAIGSFERFVILIEH
Pf'R'	471	LRDAIGREWQVGTQDYLHMPDRLGAEYVDEHSQ--RRKPVMLHRAIVGSMERFLGILIEH

1QF6	529	FAGFFPTWLAPVQAVIMNT--TDSQSEYVNELTQKLSNAGIRVKADLR--NEKIGFKIREHT
1NYQ	535	TKGAFPTWLAPKQVQLIPVNVLDHYDYARQLQDELKSGVRVSIIDR--NEKMGYKIREAQ
4P3N	300	YEGKTFPWLSPROVNVVVPV--GPTCDIYAQKVRQGFHDAKFMADIDLPGCTLNKKIRNAQ
39502'S'	528	YEGAFPAWLAPTQAVIMNT--TDKQADFAAEVEKTLNESGFRAKSDLR--NEKIGFKIREHT
obaO	530	HAGQFPLWLAPVQAVVTV--TDAQNDYADQTRNDLVQLGFRVEADLR--NEKIGYKIREST
Cshin'S'	528	FAGSFPWLAPVQAVIMNT--SESQRYTDQVAQTLRQGGLRVDADLR--NEKITYKIREHS
Cshin'R'	530	HAGYFPWLAPVQAVIMNV--TDAQADYVEAVRTALTREGFRVESDLR--NEKIGYKIREHT
Bdiff'S'	528	HAGAMPVWLAPVQAVIMNT--AESQAEYAQSLAQTLOKQGVVAAADLR--NEKISYKIREHT
Bdiff'R'	530	HAGQLEPWLAPVQAVIMNV--TDAQRYVHDVRRALIGHGVRVDVDR--NEKIGYKIREHV
Pf'S'	528	YEGAFPAWLAPTQAVIMNT--TDKQADFAAEVEKTLNESGFRAKSDLR--NEKIGFKIREHT
Pf'R'	530	HAGQFPLWLAPVQAVVTV--TDAQNDYADQTRNDLVQLGFRVEADLR--NEKIGYKIREST

```

1QF6      587  LRFVPPYMLVCGDKVEVSSGKVAVRIRRGKDLGSMDVNEVIEKLLQCEIRSRSLKQLEE-
1NYQ      594  MOKIPYQIVVGDKEVENNQVNVROYGSQDQETVEKDEH IWNLVDEIRLKKHR-----
4P3N      359  LAQYNEILVVGEKEKISGTVNIRTRDNKVVHGERTISETIERLQQLKEFRSKQAESEEF
39502'S'  586  LLKVPYLLVIGDREVEMQTVAVRIRREGADLGSMPTVAQFAEFLAQAVSRGRPDSE--
obaO      588  LQKVPYLLVVGEREKENGTVTVRSRAGEDDLGSMTMEALHAFLLNEQSAGG-----
Cshin'S'  586  LQKIPYQLIVGEEKAAAGLVAVRIRSGEDLGQMTVQAEFLERVQAEHPGT-----
Cshin'R'  588  LQKIPYLLVIGDREVEHGTVTVRSRAGDDLGTMTPAEFAARLREETAIG-----
Bdiff'S'  586  LKVPYLLVVGDKERDAQTVAVRARGGVDLGVMPTVEAFVERLQEDLRSFK-----
Bdiff'R'  588  LQKVPYLLVAGEREKACVVSVRAHSGEDLGTMTLEAEFAAHRREERPA-----
Pf'S'     586  LLKVPYLLVIGDREVEMQTVAVRIRREGADLGSMPTVAQFAEFLAQAVSRGRPDSE--
Pf'R'     588  LQKVPYLLVVGEREKENGTVTVRSRAGEDDLGSMTMEALHAFLLNEQSAGG-----

```

Supplementary Figure 6. Amino acid sequence alignment of authentic and putative ThrRSs with putative obafluorin (1) resistance proteins. The ThrRS of *E. coli* ThrRS (1QF6) was aligned (ClustalX2) with that of *Staphylococcus aureus* (1NYQ), *Homo sapiens* (4P3N), *P. fluorescens* ATCC 39502 (non-1 BGC copy – 39502'S'), ObaO, the *Chitiniphilus shinanonensis* DSM 23277 putative 1-'sensitive' ThrRS (Cshin 'S' - WP_026263085.1) and ObaO homologue (Cshin'R' - WP_026263187.1), the *Burkholderia diffusa* INT4-BP16 putative 1-'sensitive' ThrRS (Bdiff'S' - WP_059467658.1) and ObaO homologue (Bdiff'R' - WP_059467191.1), and the *P. fluorescens* sp. 34 E 7 putative 1-'sensitive' ThrRS (Pf'S' - WP_065952948) and ObaO homologue (Pf'R' - WP_065936865.1). ▼ indicates the key active site residues for substrate binding identified in *E. coli* (Figure 8.5) and ▲ represents the active site cysteine residue that is proposed to be the target of 1.

1QF6 1 ---MPVITLPGDSQRHVDHAVSPMDVALDTGP---GLAKACTIAGRVNGELVDACDLIEND
 ObaO 1 ---MVTIALPDGSRRDPEALTVQQIAQSIGA---GLAAATIIGKVDGTIVDASYLIETD
 BorO 1 MSVLRPTAETERAVVVFAGTTCADAVTAAKLPRNGPNAIVVVRDPSGAIIRDLDWTPDSD
 OzmT 1
 ThrS 1 MSDMVKITTFPDGAVKEFAKGTTTEDIAASTISPP---GLKKKSLIAGKINGKELDIRTPINED
 ThrZ 1 MSKHVHIQLPDGQIQEYPKGITTIKEAAGSISSS---SLQKAAAGQVNGKLVLDLSFKLEED

1QF6 55 AQLSITITAKDEEGLEIIRHSCAHLLEHALKQLWPHTKMAIGPVIDNGFYVDVLDRLTLTQ
 ObaO 55 ATVEIVTKSPQALELIRHSTAHLMAQAVQRLYPGTQVTIGPVIDNGFYDFVAVRPFITM
 BorO 61 VEVEAVALSSEDGLTVIRHSTAHLVAQAVQQLWPEARIIGGPIIENGFYDFDVERPFQF
 OzmT 1 -----MSRRAFRCLFEP
 ThrS 58 GTVEIITEGSEEGLOIMRHSAAHLLAQAIKRIYKDVKFGVGPVIENGFYDVEDEAITP
 ThrZ 58 AELSIVTLDSQEGLOVIRHITAHVLAQAVKRLYGEVSLGVGPVILDGFYDVKIKGKSLAS

1QF6 115 EDLEAEKRMHEIAEKNYDVIKKKVS-WHEARETFANRGEYSYKVSILDEN-----I
 ObaO 115 DDLPLTEAEMTRIVKEQLPVTRTQLP-RDEALAFFEQIGESYKTIIDA-----I
 BorO 121 EDLERVEQRMKEIISKGQRFQRREFPDRFAARAELAK--EFPYKLELVLDKGDVDAEAME
 OzmT 12 K-----PRGE---
 ThrS 118 EDLPKTEKEMKKTIVNANLPIVRKEVSR-REEAKARFAEIGDILKLEILDA-----I
 ThrZ 118 GDLEAEKEMKNIINENLEIKRIEVS-YEEAEELFAOKDERIKLEILKD-----I

1QF6 165 AHDKPGLYFHEE-----YVDVCRGPHVNMRFCHFFKLMKTAGAYWRGDSNNKMLQR
 ObaO 164 PAGETLSLYTQGE-----FTDLCRGPHVNTAKIGAFKLMKVAGAYWRGDSNNIMLSR
 BorO 179 VCGSDLTITDNLDAARTGDVCSWDLRCRPHIPSTRILPAFKLLRNAAYWRGSEKPNQQLQR
 OzmT 17 -----CGASSCAHSELMVR-----
 ThrS 167 PEGETVSIYEQGE-----FFDLCRGVHVPSTGKIIEFKLLSLAGAYWRGDSKNQMLQR
 ThrZ 167 PRGEDITLYQGE-----FVDLCRGPHIPSTGMIAFKLITRVSGAYWRGDSKNEVLQR

1QF6 218 IYGTAWADKKALNAYLQRLLEEAAKRDHRKIKGQLDLIHM-QEEAPGMVFWHNDGWTFIFRE
 ObaO 217 IYGTAWGNEKELKAYLNQLQEAERDRHKLAKQFDLHFQ-QEEAPGMVFWHPKGWSLWQT
 BorO 239 IYGTAWPTREDELKSHLAALLEEAAKRDHRRIIGEELDLFAFNKEIGRGLPLWLNCAIIRDE
 OzmT 31 -----KETTMDHRRILGRELELFDTDPLIGAGLPWLPDGAAVRHA
 ThrS 220 VYGTAFFKKADLEEHLRMLLEEAKERDRHKLKGLKLFANSQKVGQGLPLWLPKGATIRRV
 ThrZ 220 VYGVAFQKKKDLDAHLHMLEEAAKRDHRKLGKQLGLFMF-SEEAPGMPFLPKGQIVRNE

1QF6 277 LEVEVRSKLKEYQYQEVHGFEMMDRVLWEKTIQHWNDYKDAMFT-TSSENREYCIKPMNCP
 ObaO 276 VEQYRRVYRDGGYREVKSPOVLDSLWKKSGHWDNYKENMEV-TESENROYALKPMNCP
 BorO 299 LEDWARKTERKLGYSRVVTPHITQEDLMYLSGHLPPYAEDLYAPIDIDGKYYLKPMMNCP
 OzmT 72 LEEYVRAAERRAGYRHVYSPVLGKRELMEISGHWSHYSIEDMFPPEVGGEQMLLPSLCP
 ThrS 280 IERYVVDKEISLGYEHVYTPVLGSKELMETSGHWDHYQEGMFPPEMDNETLVLRPMNCP
 ThrZ 279 LERSRELQTNAGYDEVRTPFMMNQRLEWQSGHWDHYRDNMYF-SEVDTRFAMKPMNCP

1QF6 336 GHVQIFNGLKSYRDLPIRMAEFGSCHRNFPSGSLHGLMRVGRFTQDDAHIFCTEEQIARD
 ObaO 335 GHIQIFKHGLRSHRELPIRYGEGFGCHRNFPSGALHGIMRVRAFTQDDCHIFCTEEQIAA
 BorO 359 HHHMVKARPHSYRDLPIKVAEYGTVYREFERSQQLHGIMRTRGFSQNDAAHICTADQAKD
 OzmT 132 HHALTYRSRPRSHRELPIRMAELGGYRAELSGVLGGLSRVRSIHLNDAHIFCTLDQAAG
 ThrS 340 HMMINKQDIHSYRELPIRMAELGTMHRYEMSGALSGLQVRGMTLNDAAHIFVRPDIKAD
 ThrZ 338 GHMLIFKNSLYSYRDLPIRMAEFGQVHRHEYSALNGMIRVRTFCQDDAHIFVREDQIES

1QF6 396 EVNGCIRLVYDMYSTFGFEKIVVKSITRPE----KRIGSDEMWDRAEADLAVALEENNIIP
 ObaO 395 EIKAFHYQAVKVYADFGFTDIAVKTALRPEP--GKRIGSDEWVWKAENILREALSECDVE
 BorO 419 QFLEVNRMHADYYRTIGISDFYMLALRDSANKDKYHDDQMWEDAERTREAMEESDIP
 OzmT 192 EARAATEMIGRAYEALGIRPARFRLSLPGPG--GKYVADPEKWRRSTALLTEVLDASGVA
 ThrS 400 EFIRTVRLIQDVYEDFGLSDYTFRLSYRDPEDTEKYFDDEMWNKAQSLKLEAMDEIGHD
 ThrZ 398 EIKEARLIDDEVYRTFGFE-YSVELSTRPE----DSLIGDDEWVWKAENILARVLEELGLS

1QF6	452	FEYQLGEGAFYGPKEETLYDCIDRANQCGTVQLDFSLPSRLSASYVGENERKVPVMIH
ObaO	453	WEELFGEGAFYSPKIETHLRDAIGRENOVGTMOVDYHMPRLGAEYVDEHSQRKPVMIH
BorO	479	FQIDLGGAAHYGPKVDFMTRAVTGREFAASTNQVDLYTPQRFGLTYHDSDGTEKPVVMIH
OzmT	250	YEAVEGEAAAFYGPKIDVQIADGQRESTLSTVQVDFHQPERFGLRYIGADGAGHRPVMIH
ThrS	460	YYEAEGEAAAFYGPKIDVQVKTALGKEETLSTVQLDFLLPERFDLTYIGEDGKQHRPVMIH
ThrZ	453	YEINEGDGAFYGPKIDFHDKDALKRSHQCATTQLDFQPEKFDLTYINEELNEKVRPVMIH
<div style="text-align: center;"> ▼ ▼▼▼ ▼ ▼ </div>		
1QF6	512	RAIIGSMERFIGILTEEFAGFFPTWLPVQVVMNNTLDSQSEYVNELTQKLSNAGIRVK
ObaO	513	RAIVGSLERFLGILIEHHAGQFPWLAPVQAIVVTVTDAQNDYADQTRNDLVOLGFRVE
BorO	539	RAPLGSHERFTAYLTEHFAGAFPVWLAPQVRIIPIVBELTDYAEFVRDMLLDADVRAD
OzmT	310	RSVIGSVERAVAHLEEHGGAFFAWLAPVQIAVLPVTGAQVPDGEAFLRCAALGLRAE
ThrS	520	RGVVSIMERFVAFLEIEHKGALPTWLPVQFQVIPVSPAVHLDYAKKVQERLQCEGLRVE
ThrZ	513	RAVFGSLDRFFGILIEHYGGAFFVWLAPVQVQIIPVS-HVHLDYCRKVQAEIKQAGIRAG
<div style="text-align: center;"> ▼ ▼ </div>		
1QF6	571	-ADLRNEKIGFKIREHTLRVPPYMLVCGDKEVESGKVAVRTRRCKDILGSMDEVNIEKLO
ObaO	572	-ADLRNEKIGYKIRESTLQRVPYLLVGEREKENGTVTVRSRAGEDIGSMMEALHAFLL
BorO	598	-VDAGDGRINAKVRAAVTRKIPLVVVGRREAEQRTVTVRDRSGEE-TPMSLEKFFVAHVT
OzmT	369	QADPGRGSLGARIR--AARLVPPYQVVIGEEAADGRVARRLDGRRVGPVPADEALARID
ThrS	580	-VDSRDEKIGYKIREAQVQKIPYMLVVGDEAEANGAVNVRKYGEQNSETISLDEFVKKAV
ThrZ	572	-IDERNEKIGYKIRESQVQKIPYVLVIGDEEEQENAVNVRRFCHQQNEHVPFQTEFKDKLV
<div style="text-align: center;"> ▼ ▼ </div>		
1QF6	630	QEIRSRSTKQLEE-----
ObaO	631	NEQSAGG-----
BorO	656	GLIRTRSDGAGHIRPLSKA
OzmT	427	ALVSAYGTGLWDR-----
ThrS	639	AEAK-----
ThrZ	631	KQVENRGM-----

Supplementary Figure 7. Sequence alignment of putative NP-resistant and *B. subtilis* ThrRS amino acid sequences. *E. coli* ThrRS (1QF6) (used as a reference sequence – 1QF6_A) was aligned (ClustalX2) with *ObaO*, *BorO* (Borrelidin resistance - CAE45679.1), *OzmT* (Putative oxazolomycin resistance - ABS90481.1), *ThrS* (CAA99608.1) and *ThrZ* (KIX83533.1), the two *B. subtilis* ThrRSs. ▼ indicates the key active site residues for substrate binding identified in *E. coli* (Figure 8.5) and ▼ represents the active site cysteine residue that is proposed to be the target of 1.

Primer	Sequence 5'-3'	Function
F27 (16S FW)	AGAGTTTGATCMTGGCTCAG	FW primer for amplifying bacterial 16S pcr product
R1492 (16S RV)	TACGGYTACCTTGTTACGACTT	RV primer for amplifying bacterial 16S pcr product
pTS1 MCS FW	AGCTTCCTAGGTAGTCAATTGGTCACCATGGTTGGCT AGCACCTCTCGAGGCATCATATG	Sense and antisense oligonucleotides for cloning an expanded MCS into pME3087.
pTS1 MCS RV	AGCTCATATGATGCCTCGAGAGGTGCTAGCCAACCAT GGTGACCAATTGACTACCTAGGA	
sacB FW	AGTCCGGAATTCGGAATACGGTTAGCCATTTG	Primers for amplifying sacB from pFLP2, for cloning into pTS1.
sacB RV	AGTCCGGAATTCGGATCGATCCTTTTTAACCC	
pJH10TS MCS FW	AATTCTAGTCAATTGGTCATTAATTAAGTGCCTAGCA CCTCTCGAGGCATCATATGGCA	Sense and antisense oligonucleotides for cloning an expanded MCS into pJH10.
pJH10TS MCS RV	AATTTGCCATATGATGCCTCGAGAGGTGCTAGCGCAG TTAATTAATGACCAATTGACTAG	
<i>orf2345</i> KOF1 FW	TCGCGGTACCGCATAGTCTAGAGCGAAGTGAAGGTC AATGTGCGC	Each pair of KOF (KnockOut Fragment) primers is designed to amplify one upstream (1) and one downstream (2) product, overlapping the N- and C-termini of each <i>oba</i> gene coding sequence. KOF1 is cloned as an <i>Xba</i> I- <i>Avr</i> II fragment, and KOF2 as an <i>Avr</i> II- <i>Bml</i> I, into the suicide vector pTS1.
<i>orf2345</i> KOF1 RV	TCGGGGATCCCCGACTCCTAGGCACGAAAACCACAT CACCCAGCG	
<i>orf2345</i> KOF2 FW	TGACGGATCCCCGTACCTAGGGCACTGGATACCGA AGTGTCTGC	
<i>orf2345</i> KOF2 RV	CATGAAGCTTCTGGACGCTAGCCGGTAAGTGTGCGC ATCTTCGC	
<i>obaC</i> KOF1 FW	TCGCGGTACCGCATAGTCTAGACTGTGAGAGCCATC ACCCATCG	
<i>obaC</i> KOF1 RV	TCGGGGATCCCCGACTCCTAGGCACGAGAATGCGTG GGTTGAGC	
<i>obaC</i> KOF2 FW	TGACGGATCCCCGTACCTAGGGCACTGCTCAACCA GATCGGC	
<i>obaC</i> KOF2 RV	CATGAAGCTTCTGGACGCTAGCGGCTGCTCGCAGTG GTAATGG	
<i>obaF</i> KOF1 FW	TCGCGGTACCGCATAGTCTAGAGGTTGAACATGACGA AGACCGGC	
<i>obaF</i> KOF1 RV	TCGGGGATCCCCGACTCCTAGGCAGTTGATTGGCCA GTAGCTGGC	
<i>obaF</i> KOF2 FW	TGACGGATCCCCGTACCTAGGGCCTGCCAGATCGA ATACGACG	
<i>obaF</i> KOF2 RV	CATGAAGCTTCTGGACGCTAGCGAAGCACTACCGG TCACATCC	
<i>obaG</i> KOF1 FW	TCGCGAGCTCGCATAGTCTAGAGACACCCTGTCGATC AAGAGCG	
<i>obaG</i> KOF1 RV	TCGGGGTACCGGACTCCTAGGGCTCAACCAATCGA CGATCTGGG	
<i>obaG</i> KOF2 FW	TCGCGAGCTCCCCGTACCTAGGGCTGGACGAGGAGTT ACTGGCG	
<i>obaG</i> KOF2 RV	CATGAAGCTTCTGGACGCTAGCGGATATGACTGGC GATGCCC	
<i>obaH</i> KOF1 FW	CATCGGGATCCCCCTAGGCTCCTCGGAGGTCAGGTTG ATGT	
<i>obaH</i> KOF1 RV	CCATGAAGCTTGCTAGCCATCGATGAAGCCTGCCAGA TC	
<i>obaH</i> KOF2 FW	CATGGGGTACCTCTAGACTGATAGGTGAGGTGTTCCG CG	
<i>obaH</i> KOF2 RV	CATCGGGATCCCCCTAGGGTGAAAGGCCAGCAGCACT TGC	
<i>obaI</i> KOF1 FW	CGGCGGTACCGATCGACCTAGGCTGCAGGGCGTGCT GAA	
<i>obaI</i> KOF1 RV	GTCCAAGCTTCATGTCGCTAGCCGGTGCACCAGGTCT GTG	

<i>obaI</i> KOF2 FW	GCCGGGTACCGCATGATCTAGAGACGTAGTGGCAATGCTCTTGGCC		
<i>obaI</i> KOF2 RV	CGGGGTACCGCATGACCTAGGGATCAGGTTTCATCTGGTCCCGTCG		
<i>obaJ</i> KOF1 FW	TCGCGGTACCGCATAGTCTAGACAACCTCACGCTCAGTTCGCGG		
<i>obaJ</i> KOF1 RV	TCGGGGATCCCCGACTCCTAGGCCAGTCCAACGGTTGCGTTGC		
<i>obaJ</i> KOF2 FW	TGACGGATCCCCGTACCTAGGCATATGGCCAGTTGC		
<i>obaJ</i> KOF2 RV	CATGAAGCTTCTGGACGCTAGCGACAGCAGAAAGGCCACTACC		
<i>obaK</i> KOF1 FW	TCGCGGTACCGCATAGTCTAGACTCAAGTCGTTGTTGCGCCAGC		
<i>obaK</i> KOF1 RV	TCGGGGATCCCCGACTCCTAGGCTCTTGAGGCGACTCGAGTTCCG		
<i>obaK</i> KOF2 FW	TGACGGATCCCCGTACCTAGGGTGAGCTTCATGGAGCTTGCC		
<i>obaK</i> KOF2 RV	CATGAAGCTTCTGGACGCTAGCCTACCGCAGCAATATCCTGCGG		
<i>obaL</i> KOF1 FW	TCGCGGTACCGCATAGTCTAGAGGCTTAACTGGCACC		
<i>obaL</i> KOF1 RV	TCGGGGATCCCCGACTCCTAGGCAATGCCTGGGCGTTGTTGTCC		
<i>obaL</i> KOF2 FW	TGACGGATCCCCGTACCTAGGCAGGATATTGCTGCGGTAGTGGC		
<i>obaL</i> KOF2 RV	CATGAAGCTTCTGGACGCTAGCCTGCGGTATTGGCAATCAGGAGG		
<i>obaM</i> KOF1 FW	CGTCGGTCTAGACAGACGAACAACGCTCACCC		
<i>obaM</i> KOF1 RV	CGTCGGCCTAGGGTCGTCTTCACTGACCAGCC		
<i>obaM</i> KOF2 FW	GCTAGTTCTAGACGTCGGCCTAGGGAAGACGCAGTCCCTTCTTCG		
<i>obaM</i> KOF2 RV	GCTAGTAAGCTTCGTGCGGCTAGCGCAGGTCGTTGTCCTTGCG		
<i>obaO</i> KOF1 FW	CGTCGGTCTAGACGCATTACGAGTTCGTCTTCCG		
<i>obaO</i> KOF1 RV	CGTCGGCCTAGGCTGTTGCACAGTCAAAGCTTCTGG		
<i>obaO</i> KOF2 FW	GCTAGTTCTAGACGTCGGCCTAGGACTTGGAAGCATGACGATGG		
<i>obaO</i> KOF2 RV	GCTAGTAAGCTTCGTGCGGCTAGCCTGCAACTTAAGCAGGTACCCG		
pJH10TS- <i>obaC</i> FW	CCTGAAGCTAGCATGCTGAATCTCAATTGCTC	Each pair of pJH10TS primers is designed to clone the entire coding sequence of each <i>oba</i> gene as either a <i>BmtI</i> - <i>KpnI</i> or <i>BmtI</i> - <i>XbaI</i> fragment into pJH10TS.	
pJH10TS- <i>obaC</i> RV	GAGTCCGGTACCTCAAGTCGTGACATAGTTGCG		
pJH10TS- <i>obaF</i> FW	CCTGAAGCTAGCATGAGCCTTCTCCCGCATC		
pJH10TS- <i>obaF</i> RV	GAGTCCGGTACCTCATTGGGTTTTCTCTGCGCTTGAGG		
pJH10TS- <i>obaG</i> FW	CCTGAAGCTAGCATGAGCAATGTCAAACAACAAACCCGCC		
pJH10TS- <i>obaG</i> RV	GAGTCCGGTACCTCATCGTTGGGCTCCCTGATACACC		
pJH10TS- <i>obaH</i> FW	CCTGAAGCTAGCATGACCAACCTCCCCTCCACG		
pJH10TS- <i>obaH</i> RV	GAGTCCCTAGAGAGGTCATGCCTTGGGCAAGTG		
pJH10TS- <i>obaI</i> FW	CCTACGGCTAGCATGTCAGCCTCATTCACTGCTG		
pJH10TS- <i>obaI</i> RV	GAGTCCCTAGATTATTGGCCAGTTGTAAGTCTGCG		
pJH10TS- <i>obaJ</i> FW	CCTGAAGCTAGCATGCAATCCCGAAAATCATGG		
pJH10TS- <i>obaJ</i> RV	GAGTCCGGTACCTCATCGTGCCTTCCCGAG		
pJH10TS- <i>obaK</i> FW	CCTGAAGCTAGCATGACCCAGGGCAAGCTGATCTACG		
pJH10TS- <i>obaK</i> RV	GAGTCCGGTACCTCATGGCTGCTCGCACTCCGTC		
pJH10TS- <i>obaL</i> FW	CCTGAAGCTAGCATGAACCGTTTCGAACATCAGTCGATTGTCG		
pJH10TS- <i>obaL</i> RV	GAGTCCCTAGAGTCACTCAAAGGTGGCCCCACCG		
pJH10TS- <i>obaM</i> FW	CCTGAAGCTAGCGTAGTGCCTATCGCC		
pJH10TS- <i>obaM</i> RV	GAGTCCGGTACCTCATCGAGACTCCCGATAGGC		
pJH10TS- <i>obaO</i> FW	CCTGAAGCTAGCATGGTCACTATCGCTTACCAGG		
pJH10TS- <i>obaO</i> RV	GAGTCCGGTACCTCAGCCGCTGCTGATTGC		
pET28a(+)- <i>obaC</i> FW	CCTGAACATATGCTGAATCTCAATTGCTC		Each pair of pET28a(+) primers is designed to clone the entire coding
pET28a(+)- <i>obaC</i> RV	GAGTCCCTCGAGTCAAGTCGTGACATAGTTGCG		
pET28a(+)- <i>obaG</i> FW	CGTAACCATATGAGCAATGTCAAACAACAAACCCG		
pET28a(+)- <i>obaG</i> RV	GAGTCCCTCGAGTCAATCGTTGGGCTCCCTGATACAC		
pET28a(+)- <i>obaH</i> FW	GTCAGCCATATGACCAACCTCCCCTCCACG		
pET28a(+)- <i>obaH</i> RV	CACCGCCTCGAGTCAATCGTTGGGCAAGTGCTGC		

		pair for pJH10TS
pET FW	TGTGAGCGGATAACAATTC	Vector-specific screening primer pair for pET28a(+)
pET RV	GCTAGTTATTGCTCAGCGGTGG	
<i>orf2345</i> screen FW	CTTATCTCTCACTGACCCACCGC	Primers for screening across the region of <i>oba</i> gene deletions
<i>orf2345</i> screen RV	CGACCCATGCAAACACAGTGGG	
<i>obaC</i> screen FW	CACCGTCTATGTGGTTGCTCGG	
<i>obaC</i> screen RV	GGGAGTGATAACGAACCACCGG	
<i>obaF</i> screen FW	CAACTGGTCTCGACCATCAGC	
<i>obaF</i> screen RV	GTTTGGCGACGTCAATCAGCAGG	
<i>obaG</i> screen FW	GATCAACTGGTGGACTTGCTCGG	
<i>obaG</i> screen RV	CACGCGCTCTACATAGGCTTGGG	
<i>obaH</i> screen FW	CTGCTGGCCTTTCACCCTAGG	
<i>obaH</i> screen RV	GGAGATGGAACCTTCGGTCGGG	
<i>obaI</i> screen FW	CCAGATGAACCTGATCCCTAGG	
<i>obaI</i> screen RV	CCGTTATTCCTGCACCACACG	
<i>obaJ</i> screen FW	CACGCCAGATGGTTTCTACCGC	
<i>obaJ</i> screen RV	GGGTGATATCGACCACCTTCGG	
<i>obaK</i> screen FW	CTGGTCAAGCATCGCTACAGCG	
<i>obaK</i> screen RV	CCTGGAGACGAAGAAGGTACCG	
<i>obaL</i> screen FW	CCATATGGCCAGTTGCTCCAGC	
<i>obaL</i> screen RV	GAAGCGTTGATCGAACGTGGCG	
<i>obaM</i> screen FW	GATATTGCTGCGGTAGTGCC	
<i>obaM</i> screen RV	CATGGAACAATCGCGAAGCG	
<i>obaO</i> screen FW	GAATCGCACATTCATGTCTACGCC	
<i>obaO</i> screen RV	GACCTGGACCATCATTGGTCAGC	

Supplementary Table 1. Oligonucleotides and primers used in this work. Restriction sites are indicated in bold and start codons are underlined.

Organism	GenBank Accession number	Taxonomy	Function	Seq ref.
<i>Pseudomonas fluorescens</i> ATCC 39502	NA	Proteobacteria	L-TTA (ObaG)	1
<i>Pseudomonas fluorescens</i> ATCC 39502	NA	Proteobacteria	L-TA	2
<i>Pseudomonas fluorescens</i> ATCC 39502	NA	Proteobacteria	L-TA	3
<i>Pseudomonas aeruginosa</i> PA01	NP_254100.1	Proteobacteria	L-TA	4
<i>Pseudomonas aeruginosa</i> PA01	NP_249593.1	Proteobacteria	L-TA	5
<i>Pseudomonas fluorescens</i> SBW25	WP_015886246.1	Proteobacteria	L-TA	6
<i>Pseudomonas fluorescens</i> SBW25	WP_015885469.1	Proteobacteria	L-TA	7
<i>Pseudomonas putida</i> KT2440	NP_742488.1	Proteobacteria	L-TA	8
<i>Pseudomonas protegens</i> Pf-5	WP_011063906.1	Proteobacteria	L-TA	9
<i>Pseudomonas protegens</i> Pf-5	WP_011062763.1	Proteobacteria	L-TA	10
<i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000	NP_790259.1	Proteobacteria	L-TA	11
<i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000	NP_791665.1	Proteobacteria	L-TA	12

<i>Escherichia coli</i> K12 substr. MG1655	P75823.1	Proteobacteria	L-TA	13
<i>Pseudomonas</i> sp. NCIMB 10558	O50584.1	Proteobacteria	L-TA	14
<i>Ruegeria pomeroyi</i> DSS-3	AAV96391.1	Proteobacteria	L-TA	15
<i>Bordetella parapertussis</i> 12822	CAE38557.1	Proteobacteria	L-TA	16
<i>Bordetella bronchiseptica</i> RB50	CAE35697.1	Proteobacteria	L-TA	17
<i>Caulobacter crescentus</i> CB15	NP_421887.1	Proteobacteria	L-TA	18
<i>Pseudomonas stutzeri</i>	WP_003292028.1	Proteobacteria	L-TA	19
<i>Desulfobulbus japonicus</i>	WP_028581067.1	Proteobacteria	L-TA	20
<i>Shewanella pealeana</i>	WP_012155441.1	Proteobacteria	L-TA	21
<i>Yersinia enterocolitica</i>	WP_005171133.1	Proteobacteria	L-TA	22
<i>Azorhizobium caulinodans</i> ORS 571	WP_012170545.1	Proteobacteria	L-TA	23
<i>Bosea</i> sp. 117	WP_029351492.1	Proteobacteria	L-TA	24
<i>Methylobacterium</i> sp. GXF4	WP_007566405.1	Proteobacteria	L-TA	25
<i>Burkholderia</i> sp. H160	WP_008921553.1	Proteobacteria	L-TA	26
<i>Geobacter brementis</i>	WP_026841789.1	Proteobacteria	L-TA	27
<i>Aeromonas jandaei</i> DK-39	O07051.1	Proteobacteria	L-TA	28
<i>Mycobacterium liflandii</i>	WP_015356249.1	Actinobacteria	L-TA	29
<i>Brevibacterium linens</i>	WP_009882170.1	Actinobacteria	L-TA	30
<i>Streptomyces hygrosopicus</i>	WP_014669227.1	Actinobacteria	L-TA	31
<i>Frankia</i> sp. CN3	WP_027140959.1	Actinobacteria	L-TA	32
<i>Micromonospora</i> sp. M42	EWM68484.1	Actinobacteria	L-TA	33
<i>Actinomadura oligospora</i>	WP_026414616.1	Actinobacteria	L-TA	34
<i>Sciscionella marina</i>	WP_020495043.1	Actinobacteria	L-TA	35
<i>Streptomyces</i> sp. SPB78	WP_009070954.1	Actinobacteria	L-TA	36
<i>Glutamicibacter arilaitensis</i>	WP_013349325.1	Actinobacteria	L-TA	37
<i>Demetria terragena</i>	WP_018158110.1	Actinobacteria	L-TA	38
<i>Streptomyces coeruleorubidus</i> NRRL 18370	ADN26256.1	Actinobacteria	L-TA	39
<i>Streptomyces</i> sp. DSM 5940	ADY76671.1	Actinobacteria	L-TA	40
<i>Streptomyces roseosporus</i> NRRL 15998	EFE75624.1	Actinobacteria	L-TA	41
<i>Streptomyces</i> sp. SS	WP_017236989.1	Actinobacteria	L-TA	42
<i>Listeria monocytogenes</i> EGD-e	3PJ0_A	Firmicutes	L-TA	43
<i>Exiguobacterium sibiricum</i> 255-15	3LWS_A	Firmicutes	L-TA	44
<i>Bacillus thuringiensis</i>	WP_000854011.1	Firmicutes	L-TA	45
<i>Bacillus cereus</i>	WP_002014149.1	Firmicutes	L-TA	46
<i>Virgibacillus halodenitrificans</i>	CDQ33605.1	Firmicutes	L-TA	47

<i>Brevibacillus massiliensis</i>	WP_019119750.1	Firmicutes	L-TA	48
<i>Planomicrobium glaciei</i> CHR43	ETP70229.1	Firmicutes	L-TA	49
<i>Thermincola potens</i>	WP_013121253.1	Firmicutes	L-TA	50
<i>Robinsoniella</i> sp. KNHs210	WP_027294918.1	Firmicutes	L-TA	51
<i>Caloramator australicus</i>	WP_008908414.1	Firmicutes	L-TA	52
<i>Desulfotomaculum nigrificans</i>	WP_051410853.1	Firmicutes	L-TA	53
<i>Caldisalinibacter kiritimatiensis</i>	EOD01392.1	Firmicutes	L-TA	54
<i>Thermotoga maritima</i> MSB8	NP_229542.1	Thermotogae	L-TA	55
<i>Capsaspora owczarzaki</i> ATCC 30864	XP_004348623.2	Filasterea	L-TA	56
<i>Phaeodactylum tricornutum</i> CCAP 1055/1	XP_002184057.1	Heterokontophyta	L-TA	57
<i>Calothrix</i> sp. PCC 7507	WP_015127083.1	Cyanobacteria	L-TA	58
<i>Nostoc</i> sp. PCC 7107	WP_015112230.1	Cyanobacteria	L-TA	59
<i>Scytonema hofmanni</i> PCC 7110	WP_017742536.1	Cyanobacteria	L-TA	60
<i>Halothece</i> sp. PCC 7418	WP_015227610.1	Cyanobacteria	L-TA	61
<i>Spirulina subsalsa</i>	WP_017307090.1	Cyanobacteria	L-TA	62
<i>Microcoleus</i> sp. PCC 7113	WP_015182894.1	Cyanobacteria	L-TA	63
<i>Trichodesmium erythraeum</i>	WP_011611295.1	Cyanobacteria	L-TA	64
<i>Rubidibacter lacunae</i>	WP_022605981.1	Cyanobacteria	L-TA	65
<i>Synechococcus</i> sp. PCC 7335	EDX85769.1	Cyanobacteria	L-TA	66
<i>Synechococcus</i> sp. PCC 7336	WP_017325528.1	Cyanobacteria	L-TA	67
<i>Leptolyngbya</i> sp. PCC 7375	WP_006513919.1	Cyanobacteria	L-TA	68
<i>Pleurocapsa</i> sp. PCC 7327	AFY78919.1	Cyanobacteria	L-TA	69
<i>Nodosilinea nodulosa</i>	WP_017302462.1	Cyanobacteria	L-TA	70
<i>Gloeobacter violaceus</i>	WP_011142473.1	Cyanobacteria	L-TA	71
<i>Leishmania major</i> Friedlin	XP_003721570.1	Euglenozoa	L-TA	72
<i>Naumovozya dairenensis</i> CBS 421	XP_003668398.1	Ascomycota	L-TA	73
<i>Kazachstania naganishii</i> CBS 8797	CCK71098.1	Ascomycota	L-TA	74
<i>Fusarium verticillioides</i> 7600	EWG42458.1	Ascomycota	L-TA	75
<i>Talaromyces marneffeii</i> ATCC 18224	XP_002151879.1	Ascomycota	L-TA	76
<i>Candida albicans</i>	O13427.1	Ascomycota	L-TA	77
<i>Saccharomyces cerevisiae</i> S288C	NP_010868.1	Ascomycota	L-TA	78
<i>Eremothecium gossypii</i> ATCC 10985	NP_985913.2	Ascomycota	L-TA	79
<i>Saccharomyces cerevisiae</i> W3031B	AAA72430.1	Ascomycota	L-TA	80

<i>Penicillium aethiopicum</i> IBT 5753	D7PHZ0.1	Ascomycota	L-TA	81
<i>Rhodotorula toruloides</i> NP11	EMS19212.1	Basidiomycota	L-TA	82
<i>Phytophthora parasitica</i>	ETM34049.1	Heterokontophyta	L-TA	83
<i>Phytophthora infestans</i> T30-4	XP_002907740.1	Heterokontophyta	L-TA	84
<i>Halalkalicoccus jeotgali</i>	WP_008418842.1	Euryarchaeota	L-TA	85
<i>Halobacterium</i> sp. DL1	WP_009487011.1	Euryarchaeota	L-TA	86
<i>Thaumarchaeota</i> archaeon SCGC AB-539-E09	EMR73220.1	Euryarchaeota	L-TA	87
<i>Natronolimnobius innermongolicus</i>	WP_007259344.1	Euryarchaeota	L-TA	88
<i>Haloferax</i> sp.	WP_004065464.1	Euryarchaeota	L-TA	89
<i>Haloferax volcanii</i> DS2	WP_004045358.1	Euryarchaeota	L-TA	90
<i>Salinarchaeum</i> sp. Harcht-Bsk1	WP_020446474.1	Euryarchaeota	L-TA	91
<i>Halorubrum lipolyticum</i>	WP_008003501.1	Euryarchaeota	L-TA	92
<i>Thermoplasmatales</i> archaeon SCGC AB-539-C06	WP_004556422.1	Euryarchaeota	L-TA	93
<i>Arabidopsis thaliana</i> Col-0	OAP13027.1	Plantae	L-TA	94
<i>Arabidopsis thaliana</i> Col-0	OAP06494.1	Plantae	L-TA	95
<i>Mus musculus</i>	NP_082195.2	Mammalia	L-TA	96
<i>Pseudomonas putida</i> 24-1	BAD91544.1	Proteobacteria	L-phenylserine aldolase	97
<i>Streptomyces cattelya</i> NRRL 8057	WP_014151017.1	Actinobacteria	L-Fluorothreonine transaldolase	98
<i>Streptosporangium amethystogenes</i> SANK 60709	BAM98975.1	Actinobacteria	L-Thr:uridine-5'-aldehyde transaldolase	99
<i>Sphaerisporangium</i> sp. SANK 60911	BAO20189.1	Actinobacteria	L-Thr:uridine-5'-aldehyde transaldolase	100
<i>Amycolatopsis</i> sp. SANK 60206	AKC92637.1	Actinobacteria	L-Thr:uridine-5'-aldehyde transaldolase	101
<i>Streptomyces</i> sp. SANK 62799	BAJ19052.1	Actinobacteria	L-Thr:uridine-5'-aldehyde transaldolase	102
<i>Streptomyces griseus</i> SANK 60196	BAI23322.1	Actinobacteria	L-Thr:uridine-5'-aldehyde transaldolase	103
<i>Pseudomonas fluorescens</i> ATCC 39502	NA	Proteobacteria	SHMT	104
<i>Pseudomonas fluorescens</i> ATCC 39502	NA	Proteobacteria	SHMT	105
<i>Pseudomonas aeruginosa</i> PA01	NP_254102.1	Proteobacteria	SHMT	106
<i>Pseudomonas aeruginosa</i> PA01	NP_251134.1	Proteobacteria	SHMT	107
<i>Pseudomonas aeruginosa</i> PA01	NP_253292.1	Proteobacteria	SHMT	108

<i>Pseudomonas fluorescens</i> SBW25	WP_015886244.1	Proteobacteria	SHMT	109
<i>Pseudomonas fluorescens</i> SBW25	WP_015885982.1	Proteobacteria	SHMT	110
<i>Pseudomonas putida</i> KT2440	NP_742832.1	Proteobacteria	SHMT	111
<i>Pseudomonas putida</i> KT2440	NP_742489.1	Proteobacteria	SHMT	112
<i>Pseudomonas protegens</i> Pf5	WP_011063571.1	Proteobacteria	SHMT	113
<i>Pseudomonas protegens</i> Pf5	WP_011063904.1	Proteobacteria	SHMT	114
<i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000	NP_792241.1	Proteobacteria	SHMT	115
<i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000	NP_794383.1	Proteobacteria	SHMT	116
<i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000	NP_790310.1	Proteobacteria	SHMT	117
<i>Escherichia coli</i> K12 substr. MG1655	NP_417046.1	Proteobacteria	SHMT	118
<i>Pseudomonas plecoglossicida</i>	WP_013970518.1	Proteobacteria	SHMT	119
<i>Psychromonas ingrahamii</i>	WP_011769815.1	Proteobacteria	SHMT	120
<i>Methylobacterium extorquens</i> AM1	AAA64456.1	Proteobacteria	SHMT	121
<i>Burkholderia pseudomallei</i>	3ECD_A	Proteobacteria	SHMT	122
<i>Salmonella typhimurium</i>	3GBX_A	Proteobacteria	SHMT	123
<i>Campylobacter jejuni</i>	3N0L_A	Proteobacteria	SHMT	124
<i>Rickettsia rickettsii</i> str. Sheila Smith	4J5U_A	Proteobacteria	SHMT	125
<i>Burkholderia cenocepacia</i> J2315	4N0W_A	Proteobacteria	SHMT	126
<i>Escherichia coli</i> SE11	B6I5C4.1	Proteobacteria	SHMT	127
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> str. 85-10	Q3BXI8.1	Proteobacteria	SHMT	128
<i>Lactobacillus acidophilus</i> NCFM	Q5FMC0.1	Proteobacteria	SHMT	129
<i>Streptomyces clavuligerus</i>	ACJ04032.1	Actinobacteria	SHMT	130
<i>Mycobacterium leprae</i> TN	NP_302318.1	Actinobacteria	SHMT	131
<i>Mycobacterium tuberculosis</i> H37Rv	CCP43846.1	Actinobacteria	SHMT	132
<i>Mycobacterium tuberculosis</i> H37Rv	CCP42793.1	Actinobacteria	SHMT	133
<i>Mycobacterium tuberculosis</i>	3H7F_A	Actinobacteria	SHMT	134
<i>Corynebacterium glutamicum</i> ATCC 13032	NP_600221.1	Actinobacteria	SHMT	135
<i>Streptomyces coelicolor</i> A3(2)	NP_629503.1	Actinobacteria	SHMT	136
<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	NP_391571.1	Firmicutes	SHMT	137
<i>Geobacillus stearothermophilus</i>	1KKJ_A	Firmicutes	SHMT	138
<i>Streptococcus thermophilus</i>	4WXG_A	Firmicutes	SHMT	139

<i>Lactococcus lactis</i> subsp. <i>lactis</i> II1403	NP_266757.1	Firmicutes	SHMT	140
<i>Candidatus Desulforudis audaxviator</i> MP105C	B1I6M4.1	Firmicutes	SHMT	141
<i>Thermoanaerobacter</i> sp. X514	B0K631.1	Firmicutes	SHMT	142
<i>Hydrogenobacter thermophilus</i> TK-6	BAI70276.1	Aquificales	SHMT	143
<i>Thermus thermophilus</i>	WP_011228734.1	Deinococcus-Thermus	SHMT	144
<i>Synechococcus elongatus</i> PCC 6301	Q5N2P9.2	Cyanobacteria	SHMT	145
<i>Thermosynechococcus elongatus</i> BP-1	Q8DH33.1	Cyanobacteria	SHMT	146
<i>Plasmodium falciparum</i> 3D7	4O6Z_A	Apicomplexa	SHMT	147
<i>Plasmodium vivax</i> Sal-1	XP_001613892.1	Apicomplexa	SHMT	148
<i>Saccharomyces cerevisiae</i> S288C	P37291.2	Ascomycota	SHMT	149
<i>Talaromyces marneffei</i> ATCC 18224	XP_002151718.1	Ascomycota	SHMT	150
<i>Aspergillus niger</i>	CAK48606.1	Ascomycota	SHMT	151
<i>Homo sapiens</i>	3OU5_A	Mammalia	SHMT	152
<i>Oryctolagus cuniculus</i>	1LS3_A	Mammalia	SHMT	153
<i>Mus musculus</i>	AAK15040.1	Mammalia	SHMT	154
<i>Homo sapiens</i>	1BJ4_A	Mammalia	SHMT	155
<i>Paracoccus</i> sp. AJ110402	BAG31000.1	Proteobacteria	MSHMT	156
<i>Aminobacter</i> sp. AJ110403	BAG31001.1	Proteobacteria	MSHMT	157
<i>Streptomyces vinaceusdrappus</i> NRRL 2363	AEF16058.1	Actinobacteria	G/SHMT	158
<i>Streptomyces</i> sp. Sp080513GE-23	BAP16701.1	Actinobacteria	A/MSHMT	159
<i>Streptomyces</i> sp. Sp080513GE-23	BAP16692.1	Actinobacteria	A/MSHMT	160
<i>Streptomyces</i> sp. SANK 60405	BAJ05887.1	Actinobacteria	SHMT	161
<i>Streptomyces</i> sp. NRRL 30471	ADZ45329.1	Actinobacteria	SHMT	162
<i>Streptomyces</i> sp. SN-1061M	ADC96660.1	Actinobacteria	SHMT	163
<i>Streptomyces</i> sp. MK730-62F2	ACQ63622.1	Actinobacteria	GHMT	164
<i>Amycolatopsis mediterranei</i> U32	WP_013222557.1	Actinobacteria	AHABS	A
<i>Actinosynnema pretiosum</i> subsp. <i>auranticum</i> ATCC 31565	AAC13997.1	Actinobacteria	AHABS	B
<i>Streptomyces leeuwenhoekii</i> C34	WP_029387788.1	Actinobacteria	AHABS	C
<i>Streptomyces</i> sp. FXJ7.023	WP_037770928.1	Actinobacteria	AHABS	D
<i>Streptomyces lavendulae</i> NRRL 2564	AAD27811.1	Actinobacteria	AHABS	E

<i>Streptomyces collinus</i> Tü1892	CAA90935.1	Actinobacteria	AHABS	F
<i>Streptomyces collinus</i> Tü1892	AAD31828.1	Actinobacteria	AHABS	G
<i>Actinosynnema pretiosum</i> subsp. <i>auranticum</i> ATCC 31565	AAC13997.1	Actinobacteria	AHABS	H
<i>Streptomyces autolyticus</i> CGMCC0516	WP_079256968.1	Actinobacteria	AHABS	I
<i>Streptomyces hygroscopicus</i> 17997	AAO61216.1	Actinobacteria	AHABS	J

Supplementary Table 2. Source organisms and GenBank Accession numbers for SHMT, L-TA, L-TTA and AHBAS amino acid sequences used in phylogenetic analyses. Seq ref. numbers refer to sequences reported in Supplementary Sequence Data 1.

Appendix 2:
Supplementary Sequence
Data

Appendix 2: Supplementary Sequence Data

Supplementary Sequence Data 1. Trimmed sequences used in the Maximum Likelihood phylogeny. Number corresponds to phylogeny reference number allocated in Supplementary Table 2 (Appendix 1).

>1

VDWLSSTLGKDHQYREDSLSLTANENYPSALVRLTSGSTAGAFYHCSFFFEVDPAGEWHFP
EPGHMNAIADQVRDLGKTLIGAQAFDWRPNNGSTAEQALMLAACKPGEFVHFAHRDGGH
FALESIAQKMGIEIFHLPVNPTSLLDIVAKLDEMVRNPHIRIVILDQSFKLRWQPLAEI
RSVLPDSCSTLYDMSHDGGLIMGGVFDSPFLSCGADIVHGNTHTKTIIPGPQKGYIGFKSAQH
PLLVDTSLWVCPHLQSNCHAEQLPPMWVAFKEMELFGRDYAAQIVSNAKTLARHLHELGL
DVTGESFGFTQTHQVHFAVGDLQKALDLCVNSLHAGGIRSTNIEIPGKPGVHGIRLVQA
MTRRGMKEKDFEVVARFIADLYFKKTEPAKVA

>2

IDLRSDTVTQPTPGMLDAMASAVSGDDVYGEDPSVNQLEAELARRLGFDAALFVPTGTMS
NLLALMAHCERGEYIVGQQAHTYKYEGGGA AVLGS IQPQPLEVQADGSLDLSQVLDIAIK
PDDFHFAFTRRLALENTMQGKVLPMTYLAKARAFTREHGLALHLDGARIYNAAVKLGVDA
REIAGHFDSVSVCLSKGLGAPVGSVLCGSTALIAKARRLRKMVGGMRQAGSLAAAGLYA
LDHQVQRLADDHANAQWLGDELQAGYTVPEPVQTNMVYVQMGDRAQALKSFAAERGIKLS
AAPRLRMVTHLDVSRQIEQVVQTFVAFSQK

>3

MTDKSQFASDNYSIGICPEAWAAMEQANQGHQRAYGDDEWTHRAADGFRNLFETDCEVFF
AFNGTAANSLALSSLCQSYHSVICSETAHVETDECGAPEFFSNGSKLLTARTENGKLTPE
SIREIALKRQDIHYPKPRVVTLTQATEVGSVYTPPEIRAI SVTCKELGLNLHMDGARFSN
ACAF LGCSPADLTWKAGVDVLCFGGTKNGMAVGEAILFFNHKLAEDFDYRCKQAGQLASK
MRFLSAPWVGLLENDAWLKARHANHCAQLLSSLVADIPGVELMFPVQANGVFLQLSEPA
IAALTAKGWRFYTFIGKGGARFMCARDTEEERVRELAADIREVMNA

>4

MTDHTQQFASDNYSIGICPEAWAAMAEANRGERAYGDDQWTARASDYFRQLFETDCEVFF
AFNGTAANSLALALCQSYHSVICSETAHVETDECGAPEFFSNGSKLLLAQTEVGKLTTPA
SIRDIALKRQDIHYPKPRVVTLTQATEVGTVYRPDELKAI SATCKELGLHLHMDGARFSN
ACAF LGCSPAELSWKAGVDVLCFGGTKNGMAVGEAILFFNRDLAEDFDYRCKQAGQLASK
MRFLAAPWVGLQDDAWLRYADHANRCARLLAELVADVPGVSLMFPVEANGVFLQLSEPA
IEALRARGWRFYTFIGEGGARFMCARDTIERVRELARDIRLVMG

>5

IDLRSDTVTQPTAGMREAAAAELGDDVYGEDPTVNRLEAELAARLGFAAALFVPTGTMS
NLLGLMAHCERGDEYIVGQQAHTYKYEGGGA AVLGS IQPQPIDGEADGSLHLDKVA AAIK
ADDFHFAFTRRLALENTMQGKVLPLDYLAARAFTTRARGLALHLDGARLYNAAVKLGVDA
SEITRHFDSVSVCLSKGLGAPVGSVLCGSVELIGKARRWRKMVGGMRQAGLLAAAGLYA
LDHQVARLADDHANAARLGDGLRELGYAVEPVQTNMVYVDVGERAVALRDFLAERGVRI S
AAARLRLVTHLDVSAESIGQVLDAAAFRRS

>6

MTDKSQFASDNYSIGICPEAWAAMEQANQGHQRAYGDDEWTHRAADGFRNLFETDCEVFF
AFNGTAANSLALSSLCQSYHSVICSETAHVETDECGAPEFFSNGSKLLTARTENGKLTPE
SIREIALKRQDIHYPKPRVVTLTQATEVGSVYTPPEIRAI SVTCKELGLNLHMDGARFSN
ACAF LGCSPADLTWKAGVDVLCFGGTKNGMAVGEAILFFNHKLAEDFDYRCKQAGQLASK
MRFLSAPWVGLLENDAWLKARHANHCAQLLSSLVADIPGVELMFPVQANGVFLQLSEPA
IAALTAKGWRFYTFIGKGGARFMCARDTEEERVRELAADIRKVM

>7

IDLRSDTVTQPTPGMLDAMATAASGDDVYGEDPSVNRLEAELAKRLGFAAALFVPTGTMS
NLLALMAHCERGEYIVGQQAHTYKYEGGGA AVLGS IQPQPLDVQADGSLDLDQVLA AAIK
PDDFHFAFTRRLALENTMQGKVLPLEYLANARAFTREHGLALHLDGARLYNAAVKLGVDA

RAIAQHFDVSVVCLSKGLGAPIGVLCSGSTALLAKARRLRKMGVGGMRQAGSLAAAGLYA
LDHQVQRLADDHANALWLDALREAGYSVEPVQTNMVVYVQIGDRADALKAFAAERGIKLS
VAPRLRMVTHMDVSRQMERVLQAFVFEFSK

>8

MTDKSQQFASDNYSIGICPEAWVAMEKANRGHDRAYGDDQWTERASEYFRNLFFETDCEVFF
AFNGTAANSLALASLCQSYHSVICSETAHVETDECGAPEFFSNGSKLLTAASVNGKLTPO
SIREVALKRQDIHYPKPRVVTITQATEVGTVYRPDELKAI SATCKELGLNLHMDGARFTN
ACAFILGCSPAELTWKAGVDVLCFGGTKNGMAVGEAILFFNRQLAEDFDYRCKQAGQLASK
MRFLSAPWVGLLEDGAWLRHGNHANHCAQLLASLVSDLPGVELMFPVEANGVFLQMPPEHA
IEALRGKGRWFYTFIGSGGARFMCSWDTEEARVRELAADIRTIIGG

>9

MTDKSQQFASDNYSIGICPEAWAAMEQANHGHERAYGDDQWTARASDHFRKLFETDCEVFF
AFNGTAANSLALSSLCQSYHSVICSETAHVETDECGAPEFFSNGSKLLIAGTQNGKLTPE
SIREIALKRQDIHYPKPRVVTLTQATEVGSVYTPPEVRAI SATCKELGLNLHMDGARFSN
ACAFILGCTPAELTWKAGVDVLCFGGTKNGMAVGEAILFFNHKLAEDFDYRCKQAGQLASK
MRFLSAPWVGLLENDAWLKARHANHCAQLLAQLVADIPGVELMFPVQANGVFLQLSEPA
IAALTAKGRWFYTFIGKGGARFMCSWDTEEDRVRELAADIRTVMSA

>10

IDLRSDTVTQPTPGMLDAMASAPTGDVVYGEDPTVNRLEGELARRLGFAAALFVPSGTMS
NLLALMAHCERGDEYIVGQQAHTYKYEGGAAVLGSIQPQPLELQADGSLDLVQVAAAIAK
PDDHFHARTRLLALENTMQGKVLPLAYLAQARAFTREHGLALHLDGARLYNAAVKLGVDA
RVIAEHFDSVSVVCLSKGLGAPVGSVLCGSVELIAKARRLRKMGVGGMRQAGGLAAAGLYA
LDHQVQRLADDHAAQLLADGLRRVGYVEPVQTNMVVYVEMGDRAEALKAFAAERGIKLS
AASRLRLVTHLDVSAGQIEHVLSTFAEFSRN

>11

MTDQSQQFASDNYSIGICPEAWAAMEQANKGHQRAYGDDQWTARASDDFRRLFETDCEVFF
AFNGTAANSLALSSLCQSYHSVICSETAHVETDECGAPEFFSNGSKLLTARSPGGKLTPE
SIREVALKRQDIHYPKPRVVTLTQATEVGGVYQPEELKAI SATCKELGLHLHMDGARFSN
ACAFILDASPAELTWKSGVDVLCFGGTKNGMAVGEAILFFDHDLAVDFDYRCKQAGQLASK
MRFLSAPWVGLLENDAWLKARHANRCAQLLADLVKDIAGVELMFPVQANGVFLQLSEPA
IAALTARGWRWFYTFINGGARFMCSWDTEEDRVRELAADIRLVMQG

>12

IDLRSDTVTLPTAGMLDAMAHAPVGDDVYGEDPTVNLEATLAERLGFDAALFVPSGTMS
NLLALMAHCERGDEYIVGQQAHTYKYEGGAAVLGSIQPQLDVQADGSLSLQQLDITIK
PDDHFHARTRLLALENTMQGKVLPLEYLAARKLTLEKGLALHLDGARLYNAAVKLGVDA
REITRHFDSVVICLSKGLGAPIGVLCSGQPLMVKARRLRKMLGGGMRQAGGLAAAGLYA
LDHQVQRLADDHANAAFLAESLSGLGYVEPVQTNMVVYVQIGDRAAALKAFCAERGI VLT
AASRLRMVTHLNVSRQAQAEQVIAAFAAFHP

>13

IDLRSDTVTRPSRAMLEAMMAAPVGDDVYGGDDPTVNALQDYAAELSGKEAAIFLPTGTQA
NLVALLSHCERGEYIVGQAAHNYLFEAGGAAVLGSIQPQPIDAADGTLPLDKVAMKIK
PDDIHFARTKLLSLENTHNGKVLPREYLKEAWEFTREERNLALHVDGARIFNAVVAAGCEL
KEITQYCDSTFICLSKGLGTPVGSLLVGNRDYIKRAIRWRKMTGGGMRQSGILAAAGIYA
LKNNVARLQEDHDNAAWMAEQLREAGADVMRQDTNMLFVRVGEENAAALGEYMKARNVLI
NASPIVRLVTHLDVSREQLAEVAHWRFLAR

>14

MTDQSQQFASDNYSIGICPEAWAAMEKANHGHERAYGDDQWTARAADHFRKLFETDCEVFF
AFNGTAANSLALSSLCQSYHSVICSETAHVETDECGAPEFFSNGSKLLTARSEGGKLTPE
SIREVALKRQDIHYPKPRVVTITQATEVGSVYRPDELKAI SATCKELGLNLHMDGARFSN
ACAFILGCTPAELTWKAGIDVLCFGGTKNGMAVGEAILFFNRKLAEDFDYRCKQAGQLASK
MRFLSAPWVGLLEDGAWLRHAHANHCAQLLSSLVADIPGVELMFPVEANGVFLQMPSEPA
LEALRNKGRWFYTFIGSGGARFMCSWDTEEARVRELAADIRAVMSA

>15

MFFASDNSGPGVHPEILAGLTDANQGYAMAYGADTQMVAQVEKIRRIFEAPDAAVYLVATG
TAANSLALATLSQPWQTI FCSPVVAHIHEDECNAPEFYSGGAKLTLVPGGDRMTPEALRAS
ITGEEITRGVHGPQRGPVSI TQVTERGGVYSIAELQALCAVAKAYGLPVHLDGARFANALV
ALNASPAEMTWKAGVDAVSFGGKNGCMGVEAVIFFDPKHAWEFELRRKRG AHLFSKHRY
LSAQM LAYLSDDLWLSARRANDNCARLAEGLRAAGAEFLHEPQANIVFASFPRAIHRRRL
MGAGASYHLWGAELEGANEDEMLACRMVCDWSIGTDQIDRFLSL

>16

IDLRS DTVTRPSAAMRQAMAAAPVGDDVMGDDPSVRR LQDEVAARAGKEAGLFFPSGTQS
NLAALMAHCARGDEYLVGQQAHTYKFE GGGAAVLGSIQPQPV DHAADGSLPLDR LAAALK
PGGDPHFARTRL LALENTFQGRVMPAGYVAQATGWAREHGLSTHLDGARVFNAAVASGVP
VQQLCPEFDTVSI CFSKGLGAPVGSVLVGSRALIEQAHRWRKMLGGGMRQAGILAAACLH
ALHHHVERLALDHDNAARLAAGLDGIDGVRVLGQHTNMVFAEFDPARCESLTTALRGQGI
LMRAVYGGPTRLVTHLDVDADGIDRVVDAVRGHCAAAR

>17

IDLRS DTVTRPSAAMRQAMAAAPVGDDVMGDDPSVRR LQDEVAARAGKEAGLFFPSGTQS
NLAALMAHCARGDEYLVGQQAHTYKFE GGGAAVLGSIQPQPV DHAADGSLPLDR LAAALK
PGGDPHFARTRL LALENTFQGRVMPAGYVAQATGWAREHGLSTHLDGARVFNAAVASGVP
VQQLCPEFDTVSI CFSKGLGAPVGSVLVGSRALIEQAHRWRKMLGGGMRQAGILAAACLH
ALHHHVERLALDHDNAARLAAGLDGIDGVRVLGQHTNMVFAEFDPARCESLTTALRGQGI
LMRAVYGGPTRLVTHLDVDADGIDRVVDAVRGHCAAAR

>18

MTQTAPRYDFASDNVAGAMPEVMEALIAANAGTASGYGDHVSRAAADRIRAALDADAQV
RFTASGTAANAFALTLLAQPEAVLAHEHAHICTDET GAPGFFGQGVGLIGLPGASGKME
LAALEAALAQPDVSYRQPAAALS LTTATEYGTVYSEDHLRALIAPVKAKGYGVHLDGARL
ANAVAGGF DLKSI AKMGVDILVMGGTKAGSTPTEAVVFLNPDHAKRLDARLKHAGQLISK
GRFLAAPWLGLLGENGQTAPWAARAHAHANAMAQKLAALMPVPIKHPVEANGIFVEMDELA
LERLRGEGWFVYRFLDGTVRFMCSWATTPEMVEDLGAALKRVA

>19

IDLRS DTVTQPTTAMREAMLGAETGDDVY AEDPTVSRLECR LAADLGFAAGL FVPSGTMS
NLLGLMAHCERGDEYIVGQQAHTYKFE GGGAAVLGSIQPQPIEMEADGTLDLARVEAAIK
PDNFHFARSRL LALENTMQGKVLPLEY LAAARELSQRHGLALHLDGARLFSAAVKLG CDA
REITRHFDTVSVCLSKGLGAPVGSVLCGSDAFIAKARRLRKMVGGGMRQAGILAAAGLYA
LEHNIKRLGDDHRRAEWLGNE LAALGF SVEPVQTNMVYVDMGDQAAALTAFC AQRGIRLT
SGSQLRLVTHMDVHDEHVPVVVS AFAEFAENMS

>20

IDLRS DTVTPPTDAMRRVMADAVVGDDVYGEDPTIRELEEYCADLLGKEAALFTTSGTQG
NLLGIMVHCQRGDEYIVGQQAHA YRYEAGGAA ILASVQPQPLNFESDGTLDLKKVTA AIK
PDDVHFARTRL LCLENT HNGMALPLGYVQEAAKVAAMNQLKLHLDGARLFNAAVFLDVPV
KQLVQGVDTVSVCLSKGLGAPVGSLLCGSYQAIHEARKWRKMLGGGLRQAGILAAAGLYA
LQQHVSRLQEDH THAALLAQGLTGI PGIEVQFGKLQTNMVF IKTSEQHARGLSQYLKEHN
ILIGTGNP IRLVTHLGISGKDIEQVIACCRAFFKN

>21

IDLRS DTVTQPCQNM RALMASAETGDDVYGEDPSVKQLESYAAELLNKQAAVFCPSGTQS
NLMGLLSHCGRGDEYIVGNVAHTYLYEGGGAAVLGSIQPQPLTLQADATMDLSELEKAIK
PKDVHYARTKLI CLENT HGGPLPLPDGYTQNVKAI AQRHGLSMHLDGARLFNAVVAGN KSA
ATLTADFDSVSI CLSKGLGAPVGSLLVGSNKFI DEARHWRKMLGGGMRQAGI IAAAGLYA
LQNNVSRLSDDHRRASELANALGQIDGVS IPLGQANTNMLYVSVSQAMRERLAERAANYN
ILLPAGEQMRLVTHLN INDDDLKAI IELFQQA

>22

IDLRS DTVTQPDAA MRQAMANA EVGDDVY GDDPTVNALEAQAARLSGKEAALFLPTGTQA

NLVALLTHCQRGEYIVGQKAHNYLYEAGGA AVLGS IQPQPIDANDDGTLPLDKVLAAIK
PDDIHFAQTRLLSLENTHS GKVLPLSYLQQAWALTREQLALHIDGARIFNAVALNVPL
SEITQYCDTLTICLSKGLGAPVGSLLCGSAEYIKRARRWRKMTGGGMRQAGILAAAGLYA
LEHNVARLKDDHDNALWLEQQLRALGVEIVAPGAQTNVLYIQSSSELAAKLGPWMRERG
LISAGPI TRMITHININRTDLEKVVVALWREFLSEQR

>23

QQSQPQQFASDNYAGICPEAWAAMEANEGSAVAYGDDDWTHRASDAFRALFQTNCEVFF
AFNGTAANSLALASLCQSYHSVICASAHVETDECGAPEFFSNGSKLLVATTDGGKLT
AVLALATGRSDIHYPKPRAVTITQPTETGQVYTLDEIAALSATCRDLGLRLHMDGARFAN
ACAALGCSPAEMTWKVGVDVLCFGGTKNGMAVGEAILFFDRALALDFDYRCKQAGQLASK
MRFLSAPWVGVLESGAWLRNAAHGNACARRLAAAIAADVSGIEPMFPVEANAVFLRAPEAV
LQGLRENGWRFYTFIGGARFMFAWDADLARVDQLAADIRAAIAA

>24

MTDAQQFASDNYAGICPEAWAAMEANRGSAPAYGDDEWTTQASDAFRALFETPCEVFFA
FNGTAANSLALASLCQSYHSVICASSAHVETDECGAPEFFSNGSKLLIVPTPDGKLT
VRALAVSRDIIHFPKPRVVTITQPTETGQVYTVVEVRALSAMCRELGLKLHMDGARFAHA
CAALGCAPADITWKAGVDVLCFGGTKNGMAVGEAVLFFDPALAQDFDYRCKQAGQLASKM
RFLSAPWVGLLGSAGAWLANARHANDCAARLAAAVAGLPEIELMFPVEANGVFLRAPEPVL
EALRAKGWRFYTFIGGARFMFAWDADPVRVDALADDIRAAACAG

>25

QHDDSQFASDNYSGICPEAWSAMEANRGHAPAYGEDAWTARAADAFRRLFETECEVFF
AFNGTAANALALALALCQSYHSVICADSAHVETDECGAPEFFSNGSKLLTARTEGGKLT
AIRGLATNRSDIHFPKPRVVTITQPTETGQVYTLAELRALSATCRDLGLALHMDGSRFAN
ACASLGCSPADMTWRAGIDVLCFGGTKNGMHAGEAVVFFDPKLAEDFGFRCKQAGQLASK
MRFLSAPWVGMLESGAWLRNGAHGNACARRFADAVAGLPGIRALFPVEANAVFLAMP
AVTMRRLRARGWRFYTFIGGARFMFAWDAETARVDALVADLRALAAE

>26

MQHAFASDNYAGICPEALDALIAANNSGHEPAYGDDSWTNQVCDRLRDLFQTDCEVFFVFN
GTAANSLALASLCQSYHSVICHELAHIETDECGGPEFFSGGSKLLTAPDIGGKLT
PDAIE SIVTRRADIHYPKPKVVTLTQSTEVGTVYRVEEVRAIAAIKARRHLKVHMDGARFANA
VAALDVHPSEITWRAGVDVLCFGGTKNGLPVGEAVVFFDRALADDFAYRLKQAGQLASKMRF
ISAPWGLLDNDVWLRNARHANAMARLLETRLQEIPGVSIMFPSESNVFAQLPPPAKA
MRARGWKFYEFIGGGCRLMCAWDTQPD TVERFAAEVRELCAA

>27

MKHHFASDNYAGICNEAWAAMEANRGMAS SYGDDYWTAEACEKIRELFETDCEVFFVFN
GTAANSLALASLCQSYHSIVCHEMAHIETDECGASEFFSNGTKVLLVPGENGKIDLD
DAVEHTIHKRTDIHYPKPKALSITQATELGTLYSVQELQAI GELAKKHSRVLHMDGARFANA
VA SLNVAPKEISWQAGVDVLTFFGGTKNGFAIGEAVVFNKELAFEDYRCKQAGQLASKMRF
LTAPWIGMLESAGAWLKNAAHANNCARLENEIRKIPQVRVMFP SQANSVFLEMAPEALEA
LRGRGWHFYTFIGGGARFMC SWDTDTAEVANLVADIKASVA

>28

IDLRSDVTQPTDAMRQCMLHAEVGDVYGEDPGVNALEAYGADLLGKEAALFVPSGTMS
NLLAVMSHCQRGEGAVLGSAAHIYRYEAQGS AVLGSVALQFPVPMQADGSLALADVRAAIA
PDDVHFTPTRLVCLNTHNGKVLPLPYLREMREL VDEHGLQLHLDGARLFNAVVASGHTV
REL VAPFDSVSI CLSKGLGAPVGSLLVGSHAFIARARRLRKMVGGGMRQAGILAQAGLFA
LQQHVVR LADDHRRARQLAEGLAALPGIRLDLAQVQTNMVFLQLTSGESAPLLAFMKARG
ILFSGYGELRLVTHLQIHDDDI EEVIDAFTEYLGA

>29

IDLRSDVTTRPSQAMLAAMTAAEVGDDVWGDDPTVLSLQVAVADRAGKEAGLFFPSGTQS
NLAAVMAHCERGDEYLVGQLAHSYKYEGGGA AVLGS IQPQPIENAPDGTLPLEKIADAIAIK
PIDNHFARTRLAL ENT INGRVLPEDYVQEAVALVRSRGLSAHLDGARVCNASVASGRPV
ADLCAPYD TVSICFSKGLGAPVGSVLVGSTLLIERAHRWRKVLGGGMRQSGILAAACLYA

LEHNVERLADDHANARHLAEGLESSIEPVKVLSHATNMVFAQFPEADRAPLEAWLKQRGIL
TQMLYASRFVTHCDVSRNDVDTFVESVGDYFAQRR

>30

VDFRSDTVTTPSAGMRRAMADAIEVGDVYDEDPTTNRLESMAAEILGKEAGLFFPSATQA
NLAGIMAHCGRGDEYIVGQTAHAYRDEGGGA AVLGSVQPQLPNEPDGIIDLDAITSAIK
PDDFHNPIIRLLTLENTFNGLPLPADYIEGATELARSHGLATHLDGARVMNAAVATGADP
SAIADRFDVSVLCLSKGLGAPVGLLVGSKELVGRAKRIRKTLGGGMRQTGVLAAGIFA
LERNVDRMREDHRAKRLAEIFADFPGLGAGEARTNMVFLSPDGVMDAFSTFFADRDI
TGSAGKLRWVTHLDIDDEAIERVRRACKEFFGR

>31

VELRSDTFTLPTARMLEEMTRAPLGDDVYGEDPTARELEELAASLLGKEDACFMPSGTMG
NLAAI LAHCPRGTKALAGAESDIYVYEAGGASYCGGVIVEPVANAADGTMALADIEAAFP
EDPEDPQFALPALLCVENPQNRCGRVLPLEYLAGVQALARARKVAVHLDGARI FNAALA
LDVPRVIAGYADSVQFCLSKGLSAPVGSMMVGSADFIRSVRIRKMLGGGMRQAGVLA
AGLVALREMTARLDGDRHRTARRLAEGLAALPGVVVEPVETNMVFRVDTPHLTQAEFIER
SWNEGVRLAELGRGRI RAATHAGVSEAGVDRALSVFRDILAPAA

>32

IDLRSDTVTRPTADMRRAMAEAEVGDVYGEDPSVRALEEHTADLLGHEAAVFPVSGTGM
NFCALRASAEIGTEIIADAEAHIVTYELGGLAALGGVQTRTLPGLAGPLDLAELAAQIRP
RTAPKTYNMVPTSVVAVENTQARAGRIWPLDRLERLEITDDAGVVLHCDGARIWNAAV
GLGVEPRRLGELFGTSLVCLSKGLGAPVGLSVGDAEHIDRARVWRKRLGGGMRQAGVLA
AAGLYALRHHLDRLADHRRAELAATLADAAPGRVEPKLVETNMI FVSVDAEDFAARA
AAAGVLVGTSGPTSLRLLTHLDVDDDIRAAGAVLARLLTA

>33

IDLRSDTVTRPTPGMREMAAAAEVGDVYGEDPTVTALEAEVAALFGHEAALFAPSGSMA
NQIALQLHVPPGDELLCDADAHVVTYEVGAAAAYGGISSRTWPAVGADLDPDVVAGMIRP
DGYFAVPTRAIAVEQTHNRGGGGVILPPLTRRMREVADEHGVALHCDGARIWHAHVADRV
PLAGYELFDTLVCLSKGLGAPVGLSVGSAEKIARARVLRKRMGGGMRQVIGILAAAGR
YALAHHVERLAEDHAKAARLAEAVAPHGVLASVVRTNIVPLDLTKHPLDAHAAAAAREA
GLLISVLGPRTARLVTHLDVTDEQIDSAATILDALRA

>34

LRVHGHFGSDNHAGVHPEVMAALAEANTGYQIAYGSDVWTERLDRVIERHFGPQATAFP
VFNGTGANVVGLQAMCERWSSVICAESAHNQDECGASEHVGGKLI PVPAENGLTPEA
VARRATGFNGVQHSQPAVVSITQVTEVGTVYTPDEIVALASCAHERGMFLHMDGARLANA
AVSLDVP LRAITTDVGVVLSFGGKNGMMLGEMVVVNLQDAAHGSAYLRKSSMQLASKM
RFLSAQFVALLDGLWRRNARHANAMARLLAEQVRNIPGVEIVRPVDANAVFARLPERAA
EHLRERFSFHGHPSEEVWRMCAFDTTEADVTAFTTALHKSMTA

>35

HDPEYRGFASDNYSVHPEIMAAALATANGGHQIAYGEDDYTEAVQQVFQSHFGPGAFAFP
VFNGTGANVVGLQSLTERWQSVICTETAHIVHVECGAPERMGGLKLLTVSTPDGKLTPEL
VDTEAYGFDDHRAQPRVVSITQSTELGTCYTPAEVAALCEHAHARGMVVHMDGARLSNA
AAHLGVPFAEFTTEAGVDVLSFGGKNGMFMGEAVVVLNHEAARGLPYLRKSMQLASKM
RFVSAQFDALLSGDLWLRSAHANAMARLAGTVRDVPGRLRTREPQANEVFAILPKDVT
ARLQKRFRFYTWDEETGEVRWVTTFDTTTEADIDTFAAAVREELQA

>36

HDPAIRGFASDNYSAGAHPEVLAALAVANGGHQVAYGGDAYTEHLQVIMRDLFGPRAEAFP
VFNGTGANVVSLQAVTDRWGAVICADSAHIVHVEGGAPERVGSLLKLLTVPTPDGKLTPEL
IDREAYGWDDEHRAMPQAVSITQSTELGTYTPDEIAAICAHHERGMKVHLDGSRLANA
AAALGVPVKAFTTDAGVDLLSFGGKNGMFMGEAVVVLDPDGI RHMKHLRKLMSQLASKM
RFVSAQFEALLADGLWLRSAHANAMAQRLLAAGVVRGIDGVEILYPVQANGVFARLPHEVT
ERLQKRYRFYIWDEAAGSVWRMCAFDTTEDDVDGFLAALREEMTR

>37

HDISVRGFASDNYSVHPRVMDAIARANGGHQIAYGEDHYTEELGNVARNVFGPEAQIFP
VFNGTGANVNTALQSLPLRWGAVICASSAHINVDENGAPERVGGFKLLQVDTDPDGKLTPEL
IDKEAWGWGDEHRAQPLAVSITQATELGTIYTPPEEIKAITDHAHKLGMNVHLDGSRLSNA
AASLNVPRLALTTDAGVDIVSLGGTKNGILLGEGILTLRSELAADLKYLRKMMNQLSSKM
RFISAQLIELYGTTELWRELASHSNAMAAKLSQAVGDIIEGVELVYPTQANGVFAQLPTEIS
DQREHFRFYDWDRAAGQVRWMCSDTTEEDVEAFISKRELCGA

>38

HDTESRGFASDNYAGIHPEILTAIERANGGHQVAYGGDDYTAALQDVIRSHFGEEAAQAFP
VFNGTGANVVALQAVTDRWDGVVCTTAAHINVDECAAPERMGGLKLLQVETSDBGKLRPDQ
LDAFALKHDEHAAQARVVSITQTTELGTCSYVEEIAAVSRRAHDLGMVVHVDGSRLSNA
AASLGVGFREMVTDGTVDLLSFGGTKNGLMVGEAVIVLNPDAVRGMKFLRKQTMQLASKM
RFISAQLMALLSDDLRLRNASHANAMAQRLAEGVRDLPGVSIARPVQSNVFPVLPHEVS
RRLMESFLFYFWDETAGVVRWMCADFDTTEADVDAFVAAVRQEIDC

>39

IEMRSDTFTLPTEQMVSAQTAVLGDVVYGEDPTANRLEELAAKSVGKPAACLMPSTMA
NLAALLVHVPRGGKVLVGNESDIYLYEAGGASVCGGIVYEPITRDPDGTALDDLAAAFP
PDPDDPQFALPGLICVENTHNRMGGRVLSQAYLAELKRFATGHGIPVHMDGARIFNAAVA
TGVAEQIAAHADSIFCLSKGLSAPIGSILAGEADFIKARRIRKMLGGGMRQAGVFAA
AGLVALTSMIDRLAEDHQRAAQLAAGLAEVDGIDVDPSSVTNIVLFRVTANGLDDDRFL
RAVEQRGLVMGEFGHGRIRAVTHRGLSSADVSAVAIVADVREAS

>40

IEMRSDTFTLPTEEMLSAMGRATLGDDVYGEDPTVNHLEELAAKSVGKPAACLMPSTMA
NLAALLVHVPRGGKVLVGNESDIYLYEAGGASVCGGMVYEPITRDPDGTALDDLAAAFP
PDPDDPQFALPGLICVENTHNRMGGRVLGQAYLAELKAAFAAVHGVPHMDGARIFNAAVA
SGLPAEEIAAHADSIFCLSKGLSAPIGSILAGSAEFIGKARRVRKMLGGGMRQAGVFAA
AGTVALTDMVDRLAEDHRLAARFAAGLAEVDGIDVAPGSVTNIVLFRVTAEGLDAARFL
EDARQRGLAMGEFGHGRIRAVTHRGLPADVPAVAIVADVRAAR

>41

IEMRSDTFTLPTEEMLSAMGRATLGDDVYGEDPTVNHLEELAAKSVGKPAACLMPSTMA
NLAALLVHVPRGGKVLVGNESDIYLYEAGGASVCGGMVYEPITRDPDGTALDDLAAAFP
PDPDDPQFALPGLICVENTHNRMGGRVLSQAYLAELKAAFAAVHGVPHMDGARIFNAAVA
SGFPAAEIAAHADSIFCLSKGLSAPIGSILAGSAEFIGKARRVRKMLGGGMRQAGVFAA
AGTVALTDMVDRLAEDHRLAARFAAGLAEVDGIDVAPGSVTNIVLFRVTAEGLDAARFL
EDARQRGLAMGEFGHGRIRAVTHRGLPADVPAVAIVADVRAAR

>42

IEMRSDTFTLPTEQMMSAMSRALGDVVYGEDPTVNGLEELAAKSPSGTMANLAALMVHV
PRGGKVLVGNESDIYLYEAGGASVCGGIMYEPITRTPDGTALDDLAAALPPDPEDSQFA
LPGLICVENTHNRMGGRPLSQEYLAELRRFATRHLPHVMDGARVFNAAVATGVPAAEIA
AHADSLQFCLSKGLSAPIGSILAGEADFIKARRVRKMLGGGMRQAGVFAAAGTVALTST
VDRLADHRLAARFAAGLAEIDGIDVDPDAVTNIIIFRVTAAPGLDDDRFLRAVQERGLA
MGEFGHGRIRAVTHRGLDAEDVSAVAIVADVREAR

>43

GMTNLTLSYQKTPYKLGNGPRNVGLTEALQNI DDNLES DIYNGAVIEDFETKIAKI
LGKQSAVFFPSGTXAQQIALRIWADRKENRRVAYHPLSHLEIHEQDGLKELQQITPLLLG
TANQLLTIDDIKSLREPSSVLIELPQREIGGQLPAFEELEKISEYCHEQGISLHLDGAR
LWEITPFYQKSAEEICALFDSVYVSFYXGIGGIAGAILAGNDDFVQEAKIWKRRYGGDLI
SLYPYILSADYYFEKRIGKXAEYFEAAKGLAERFNCSGKTVPEVPVSNXFHVYFENSA
DEIGAILTKIQDETGVGISGYLQEKSAADVCAFEVSVGDFAEIPAKNLELVFRC

>44

GMNRLRTSFQQTGGQISGHGKRNVGLKTAFAAFAVDEMASDQYGTGAIIEPFQKQFADVL
GMDDAVFFPSGTMAQQVALRIWSDETDNRTVAYHPLCHLEIHEQDGLKELHPIETILVGA

ADRLMTLDEIKALPDIACLLLELPQREIGGVAPAFSELETISRYCRERGI RLHLDGARLF
EMLPYYEKTA AEIAGLFD SIYI SFYKGLGGIAGAILAGPAAFCQTARIWKRRYGGDLISL
YPYIVSADY YELR KDRMGQYYEQAKQLAEQFNALPGVHTTPEVPVSNMFHLHFDGQAAD
ISP KLEQVQEETGLGFVGYLVDKDGYCSTEISVGDAYGELDQQTRDAGFAR

>45

INYLSDTITLPT EEMLKAIQNAELGDDVYGQDKTVNELEGYAANMLGMEAA CFMPSGTMA
NLASILAHCPRGSTVLVGD ESDIYI YEAGGASICGGIMYQPIKTQPDGRLLLS DLEAAFP
NDVKDPQFALPALICLENTHNRCGGKVLPLKYLKEIKEFAMRRNIPVHMDGARVFNAAY
LGIEVKEIAQYVDSIQFCLSKGLSAPIGSMVVGSRKYIEEVRRIRKMLGGGMRQVGI IAA
PAMVALQTM TGR LQDDHIRAKRLAIGLSSIDGIEVDASEIQTNIVMFRI TDPNYDWKFL
KKTEEKGLV FSEMGYGRMRAVVHRHIKESDIEHTIKI IADLLNKSS

>46

IELRSDTFTLPTVEMLKAI IHAPLGDDVYQEDPTVNELETLVAKMMGKEAA ILMPSGTMA
NLAALMAYCPRGSKVLVGN ETDIYI YEAGGASVCGGIMYEP IPTQADGRLLIKDLARAFP
EDPTDPQFSLPSLICLENPHNRMGGRVLP LSYLKEVQMF SRQKEIPIHMDGARI FNAAIA
MDIPVKEIAQYADSVQFCLSKGLSAPIGSMIAGKKDFIQKVYRIRKMLGGGMRQAGI IAA
PALIALKQMNDR LTEDHVHARQLAEGLAQIEGIECDVDSVDTNIVFFQVIDKRYTWKTFI
EKAREHDLNIAELGHGRIRAVTHSGVNTQDINQALKI IKQIMQDNN

>47

SDCTPRFLASDN TSGICPEALDYLLRNASDD LAYGNDAWTQKAADRFRFFDYDCDVFF
VFNGTAA NSLTLASMGQSYHSVICHEL AH IETDECGGPEFFSNGAKLLTCSGADGKLTPE
GIESLVTRRS DIHYPKPRTISLTQATEVGT VYTRDELMAIRAMADKHGLHVHMDGARFAN
ACASLEATPAELTWQAGVDALCFSGTKNGLPMGEAVVFNHALAEDFGYRCKQAGQLSSK
MRFITAPWLG LLESGAWLRNAEHANGMARYLAAGFAALPGAELMFPAQANSV FVTLPVAV
LETLRQRGWTFFYTFIGAGGARFVCAWNTTQALLDRLLDDARDALSA

>48

MRKSFASDNYS GIHPEVLESIRANTGHVPSY GKD PYSEAA I LKCKEHFGENIEVFFVFN
GTGANVLGLQAITR PYQSI ICSDQAHLNVDECGAPERFTGCKLLTVPTADGKITPDQIEQ
HLIGFGDQHHSQPKV VSLTQSTEFGT VYRPEEIKAIADLAHEHGLLLHMDGSRLANAAAY
LGVPLKALTADVGDVLSFGGTKNGLMIGEAVIFFDPDLARDFMYIRKQGMQLGSKLRFI
AAQFEALLSNDLWLQNARHSNRMASLLASRIEKIPGVRITQKVEANAVFAVIPRECIATL
SERYYFHVWSERESLVRWMTAFDTTEDEIDQFAEDIAQALSS

>49

MTHKMFASDN NSGIHPAILEAIAKANS GHAPAYGSDSYTADAIRKFKEHFSEDCEVALVF
NGTGANVTGLKAMVRSHQAI ICTDGAHIHADEGGAAEGFI GCKLLTVPSADGKLTVDGIA
EQLHHIGNQORSQPKAVSISQCTELGT VYTI EEI KEITSF AHANGLYVHMDGARLSNAAA
HLGCSFKEMTV DAGIDMLAFGGTKNGLMLGEAVVCFNKELASELTYIRKQGMQLGSKMRF
IAAQFERLLTDDLW HKNASQANRMAQLLAAGLQKFPDAALTQPVESNAVFISLPKEQIGQ
LQKTYAFALWNA AINEIRLVTSFDITEEDIKEFLSHVEKAYS

>50

MISIKISKNFASDNCSGVHPEILQAI CEANQGHTLAYGDDVYTASALS K FREHF GKDIEV
FFVFNGTGANVLSLQAATDSFNAVICADTAHINCDECGAPEKFTGCKLLTIP SANGKITV
DQVEKFLHLAGNQHHSQPKAISITQSTELGT VYQPEEIQSIADFAH THDMFLHVDGTRLA
NAAASLNLTLREISR DVGVDILSFGGTKN GMMYGEAVVIFNKNLHKNFSYIRKQGMQLAS
KMRFIAAQFEALLSNDLWLRNARHTNKLAKLLAEVSKIPQIKITQPVQANAVFAIVPPQ
FIPALQQEYFFYVWNKETSEVRW MISFDTTEEDVMDVRLIRQTIS

>51

MNGKRGFESDNCS SVHPRVMQALMDVNC GHVPGYGYDPVTAEADRI FNHLFDRETDVFFT
FNGTGTNCAALAH LVTWQSILCADSAHINSAETGAPERIANAKLVPVPSIDGKISPRSL
EDAVGGWQSEHV PKPRVLSLTQVTDGTVYTPGELKNLCKIAHDHNMIVHMDGARLANAV
AANQNDLTGTTWAGVDVLSFGGTKNGLMFGAEAVVFDHKLAENFKYTRKSCGQLPSKMR
YIAAQFIEILKDG LWLE MAGHANRMMMLYEKMSGLPVFRTPFVPPQANELFAHVDERIKW

QLCDAFFPQHFGPEPGISRFBVTSFDTTEEDLNALIACAEDLTHS

>52

MRSFASDNNARVHHEILEAIIIRANDGDVVISYGDVVYTRAIIEKFKEIFGKDIDVYFVYNG
TGANVLALTAITKPFNAIIASPMSHINVDEC GAVEKSSGCKIITVPSSDGKIRIDDIKKL
LHAKGVEHHSQPKVISITQSTEVGTVYTVVEEIKEITKFAHENDMYVHMDGARIANAAASL
NKSLREITKDAGVDVLSFGGTKNGMMFGEAVVFFNRELSKDFKFKVRKQGTQLHSMRYIS
AQFEALLSNDLWLKNAKHANEMAQYLEEKVKEIDEIKLTQRVEANAVFAIMPRKAIKELQ
EKYFFYIWDEEKDEVRWMTSFDTKKEDIDL FVEEIKRAMRD

>53

VDLRSDTITLPTDEMRAAMAKAVVGDDVYGEDPTVNQLEQEAALVGKEAALFVPSGTMG
NQIAVLTHTGRGDEVILDSEAHIIYYEYVGGPAMLAGVQLQPVNGLLSDKGPELLKAALRS
EDIHFPHSTLVCLNFTNRGGTILPPNIMSEIYAIKQRGLQVHVDGARI FNAAVSLGI
DVREFTQYCD SVMFCLSKGLAAPVGSLLAGSKDFITRARKYRKALGGGMRQAGILAAAGL
VALKSIDRLAEDHANAKRLAAGLAELPGLHIDPARVQTNIVIVEVKGRLTAAELVNQLAE
RGVKCSTFGPNLIRLVTHKDVAAEDIEYALAAAKEILC

>54

VDFRSDTITKPTTEEMRKAMYEAEVGDVVYKEDPTIKKLERVAADIVGKEAALFVPSGTMG
NQLAVLTHTERGQEVILEDWSHICRFEVGGIAFLSGVQAKTVKGRVMSLKD VNEAIVL
DNDIHHPTGLICLENTHMAGGVVPI DVMNNIYNLAKENKLPVHLDGARVFNAAYV LG
CEVKEITKYTDSVMFCLSKGLSAPVGSILAGSEAFVEKARRYRKMFGGMRQAGVIAAAG
LVALDKMIDRLKEDHDNIKLLASKLNQIDGIEIDNESVQTNILMINIESSKYSSKELVEE
MKKRGILTSVITDSIIRFVTHRHITKDNINYAVDKINEIMN

>55

IDLRSDTVTKPTTEEMRKAMAQAEVGDVVYKEDPTINELERLAAETFGKEAALFVPSGTMG
NQVSI MAHTQRGDEVILEADSHIFWYEVGAMAVLSGVMHPVPGKNGAMPDDVRKAIRP
RNIHFPRSLIAIENTHNRSGRVVPLENIKEICTIAKEHGINVHIDGARI FNASIASGV
PVKEYAGYAD SVMFCLSKGLCAPVGSVVVGD RDFIERARKARKMLGGGMRQAGV LAAAGI
IALTKMVDRLKEDHENARFLALKLKEIGYSVNPEDVKTNMVILRTDNLKVN AHGFIEALR
NSGVLNAVSDTEIRLVTHKDVSRNDIEEALNIFEKLFKFKFS

>56

DTRSDTIVTKPTPGMLQAMLNAAVGDVVYDEDPSIHKLEAFVAQLTGKEAALFCATGTMTN
QLAIRSQIGALES LICDIRAHIFRYECGGVAYHSQAHIIPLQPPNGGNITLEMIEAQVLT
FNVHHA VTKLIEIENTLGGAVFPYDEIVRISTYAKQHGIKMHLDGARIWNAS IATGIPLK
QYCDHFDSVSLCLSKGLGAPIGSVLVGSRELIHRARHFRKLFGGGWRQAGLLAAGGQYAI
EHQWTRMADHDHANAKQFAELITKAGHMTVTNKVETNMVFLNPELRKKGVSWNKIAARL
RSVHGIVVNEDGAHPAPVRFVFLHITSEATAKLAAAINDEIAQ

>57

IDLRSDTVTKPSRPMLEAALSARTGDVVMGEDPTVLELEETVADLFRKEKGLYVPTGTMS
NLVAIMSHCNRRSSEIIIGANCHICLWEGGNTAGLGGVHTRQLLEDETTAQIHRDHISDA
FRSDDDDHFAKTELLCLEN SHMMGGVALPLSYVEDIGHLTTELGIKHLIDGARI FNASV
CHNVFVDNLCKPADSVSVCLSKGLGAPLGSVLVGD TDFIRLAKRARKRCGGGMRQAGVVA
AMGLYAIHNNISRLEADHRRAKRLANELQKNGFY LARNGAVDTNMFYFGLPDNSKVEWKD
YCKILQTKHGVLGTGGYSRGGRLFRVAVTHL DISDKDIDRAVEAMVLAARLS

>58

IRSHSEQFASDNSSGICPEAWEYMLRANQGSAPAYGNDEWTQKAADYFRELFEIDCEVFF
AFNGTAANSLSLAALCQSYHSVICHE TSHIETDECGAPEFASNGSKLLAKGENGKLT PQ
SIEAIVTRRADIHYPKPKVISITQATELGTLYTPEELWEIKAVAQKYNLKIHM DGARFAN
AVAALNKSPA EISWKSVDVLCFCGTKNGMALGEA I LFFNRALAEFDYRCKQAGQLASK
MRFISAPWLGLETGAWLKNARHANQCAEYLENKLLNITGVEMMFPREANAVFAKLPEHV
IHSLKAKNQOYFTFIVGGVRFMCSWNTTQSRIDELIGDIAAIA

>59

MSMNAEQFASDNYSGICPEALEYMLQANLGSAPAYGNDEWTQKVSDYFRELFEINCEVFF

TFNGTAANSLSLAALCQSYHSVICHEHAHETDECGAPEFASNGSKLLLAPGENGLTAQ
SIEAIVTKRQDIHYPKPKVISITQSTELGTVYSIAELLEIKEVAQKYKLIHMDGARFAN
AVAAMNKTPAEITWKSVDVLCFCGKNGMALGEAILFFNKALAEDFDYRCKQAGQLASK
MRFISAPWLGLLETGAWLKNARHANQCAEYLENQLLQIDGVEIMFTREANAVFVKLPEQA
IQKLDKNWQFYTFIGVGGVRFMCSWNTTKTRMDELVRDIQAAIA

>60

MDSNLEQFASDNSSGICPEALEYTIKANQGSVPAYGNDIWTQKATDSFRELFEIDCEIFF
VFNGTAANSLSLALCQSYHSVICHEHAHETDECGAPEFASNGSKLLLAKGQNGKLTDP
AIEEIIINKRADIHYPKPKVISLTQATELGTIYSIDELIAIKSMAKKCNLKIHMDGARFAN
AVVSINKSPAELTWKSVDVLCFCGKNGMAMGEAIIFFNQALAEDFAYRCKQAGQLASK
MRFIAAPWLGLLETGAWLKNALHANQCAEYLENQLLQLEGEVIMFPREANSVFVKLPELA
IAQLREKNWQFYTFIGVGGVRFMCSWNTTQARMDELVGDIKEAIGR

>61

MLEQFASDNYSGICPEVMAYMNRANQNAVAYGNDEWTQKAADYFRTLFTDCDVFVFN
GTAANSLSLATLQCQSYHSVICHEVAHIETDECGAPELASHGSKLLLGRGENGLTPESIA
EIVNKRVDIHYPKPRAISITQPTVEGTLYSIEELMAIQDVAQQYGLKIHMDGARFANAIV
AMNQTPADLSWKCGVDVLCFCGKNGMAVGEAIIFFNKSLAEDFAYRCKQAGQLASKMRF
IAAPFLGLLETGAWRKNANQHANCAEYLEERLSVIPEIEFMFPREVNSVFVKMPERVIQS
LREKSWHFYTFIGVGGVRFVCAWDTTKARIDELVNDIKDAIAI

>62

MLEQLEQFASDNYSGMCPEVLDALMSANQGSPPAYGNDQWTEKAANAFRELFEIDCEVFF
VFNGTAANSLSLALCQSYHSVIAHEVAHIETDECGAPEFFSNGSKLLLSSGENGLTPE
GIAKIVTKRQDIHFPPKPKVISLTQATEVGTVYQIEELQAIKAVADQYQLKIHMDGARFAN
AVVSLGKSPAELTWRSGVDVLCFSGTKNGMAMGESVIFDRTLAEDFAYRCKQAGQLASK
MRFIAAPWLGLLNRDWTWLNARHANCAAYLASQLAAIPELTLMPFTEVNGVFVEMPEGL
IQALREKGLWLFYNFIVGGIRLMCSWATTRKRIDELVADIKSAIIS

>63

MNFASDNVTGISPEIMAALSAANAGVAMPYQDEYTRQLKTQFNDFETDVTVPVATGS
AANALALSVMTPPFGAIYCHTESHINVDECGAPEFYTGAKLVTLPGSQKILATDLATV
LAQAGAGVVHHVQPAAVSLTQATEAGTVYTPDEIGALAEVAKSHRLYLHIDGARFANA
SLGCSPADLTWRAGVDVLCFGATKNGAMAAEAVVFFNQTLAETFGYRRKRSGHLFSKMRF
LSAQLLAYIEQDLWLNANHANQMAQKLAQGLVELPGVELCYPVQANEIFIQLPKALISS
LLADGFYFYPWGENSTTVRLVTAFTNTSEADVMAFIEAARRYSSE

>64

MNFCSDNVAEVCPEIMAALIAANEGAMMPYGEDEYTRQLEAKFSSIFETPVTVPVATGS
AANALSLSAIAPGYGAIYCHSESHINVDECGAPEFYTGAKLVTLNGTDAQINPSELSKT
LEKAGAGVVHHVQPAAVSITQATEAGTVYLPPEEIGEIAKITHAHNLYLHIDGARFANAVV
SLDCAPADITWRAGVDVLCFGATKNGAMAAEAVIFFNQDLAKTFAYRRKRSGNLFKSMRY
LSSQLEAYITDDLWLNANFANQMATRLAQGLVDLPLVKLCHPVEANEVFAEIPKSVNG
LRGDGFQFYVWEEGAFVPIRLVTAFTNTKKEDVDALLKSIQRHLRV

>65

LDFRSDVTVKPLRMREAMFKAIEVGGDDYRDDPTVQELERFAADLFGKEAALLTSSGTQS
NLIAIMSHCQRGEAFIVGHRSHAYLRELGGAAIAGVQPQVHNRADGKLSLEDIKASLF
PASILFAPVRLIAVENTIDGMVPLDYLDKDVADLATEHGLQVHTDGARIFNAAAIGCPV
SNIACFSNTVSFCLSKGLGAPVGSVLVGSQAQTIKARRHRQMLGGGMRQAGILAAAGLFA
LQNNVSRLERDHARAKELADALSINLPLGTVEVTPQTNIVFCNVDAVQKGFSEFLKKNIGR
VSGTTPDRQRWVTHLGI DDVALATTANLLTQMGN

>66

IDLRSDTVTOPTPEMRQVMLAAPVGGDDVLGDDPSVNELETYVAALLGKKAAYMPSGTMT
NQVALRSHTOPGDEIVLEAEAHIIYYEGGGPAALSGVSCRLLSGEGGIFSAAAVEKSLRP
VDPHFPRTRLVCVENTHNRSGGRVFPLETIQEIAQVCQRNDLQLHMDGARFVNACVATGL
SPQEYAAPFDTVSVCFSKGLGAPVGSALVGSQAQKIATARRFRKMFGGMRQAGMIAAGAL

YAVQHQFERLAEDHYNAQRLAKKLGKIGKISVDPALVETNIVNFELTKFSASDVIKTLAK
EGVAVLATGPRKIRAVTNLMVSTEQIDQAVEAIARSVW

>67

IDLRSDTITQPTPAMRLAIAANAAGDDVLGDDPTVKELEADVAQLLGKEAAVYMPSTMT
NQVAVRLHTQPGDEIVLESEAHIIYYESEGAPAALSGVMCRLVSGRQGVFAADDLVNVLRS
GNPHYPTTLVCIENHNRRGGTIFPLEAIQAIAEVCQQRGLAIHMDGARLWNACEATGI
TPTDYAEFPDTSVCFKGLGAPVGSALAGSSETIDRARRFRKMFGGMRQAGTIAAGAL
YALQHHDRLEDRENAQLLARQLVEVDGIEVDLSTVQTNIVVFRCLDVPAPLLAQALGD
RGVKLFARDRHSLRAVTSMLVTRHEILQVKDKVADALTEVC

>68

IDLRSDTITQPTRGMREAIATAEVDGDDVLGDDPTVQQLERFVAGLLGKEAAVYMPSTMT
NQIALRTHTEPGDEVILESEAHIIYYEGGPAALSGVMCRLINGDRGVFTAADVETVLRP
ADVHFPKTKLVCLNHNRRGGKIYPLSEIDATAQVCKQHGLKHLHDGARLWNACIATNT
PEDKYVKAFDVSVCFSKGLGAPIGSALAGSHDFIERARRFRKMFGGMRQAGMMAAGAI
YALKHHRERLLEDHNAQTAKGLAEIPGIIIDPIAVETNIVVVFHTKMLRAQVLVDRKLK
KGVYVLPVGPKSIRAVTNLMVTEEQILQVPDLVDSLNES

>69

IDLRSDTITKPTPAMREAMAKAEVGDVDFGDDPTVNALEAYVAELLGKEAAVYMPSTMT
NQVALRVHTEPGDEVILESEAHIIYYEGGPAALSGVMCRLIKGNKGI FSATDVERVLRP
DNYHFPKTKLVCLNHNRRGGRIYPLSEIEAIAHICQERNLKLHLDGARLWNACAATSI
SEADYAKFPDTSVCFKGLGAPVGSALVGSKELIKRRARRFRKMFGGMRQAGIIAAGAL
YGLKHHRRERLLEDHVNAILAKGLQQIDGIAIDPEDVQTNIVIFQTKAIPAETLAQNLQE
KGVALLAIGSHSLRAVTNLMVTQEIQAVPELVEAMRS

>70

DFRSDTVTWPTPAMVEAMATAEVDVYGEDPTVNELESLAASLLGKAAGLFVTSQTQGN
LIAALTHAQRGDEAILGEDAHTFCWEAGGIAVLGSIITRPLPTDSRGRMDVDRIKASLRV
DNPPLPHSRLILLENSSGGNGAAIEPSYFESIGALAKQHDLAVHLDGARLFNAATALRL
DPTMITAHVDSVSVCLSKGLCAPVGSVLVSEAFIHQARRHRKLLGGGMRQAGVLAAGI
IALKTMSQLQIDHDNAQTLAQGLATIPGIIIDPSVETNMVFFDLAPEVAIAPEELIKA
LRTDYGLYMGYGDRTLRVAVTHYWIEQPQVERLVEAIRTILTGAA

>71

VDLRSDTVTRPTRAMRAAMFEAEVGDVYGEDPTVNRLQERVAHLLGKEKGLFFPSGTMC
NQVAIATHTRPGQEIICDRDCHVFNYERAGASHLGGVQLHPLDGRHGVLSAQQVRSVAVRP
IELHEPPTGLILLENHNRRGGTVYPQDVIEQIAAVAGAAGIPMHLDGARLLNASVAGSH
SPATLAAPFDVSVYFCLSKGLGCPAGSVLAGKADFAQATRLRKAFFGGMRQVGFLLAAGL
YALDHHIERLADHRRARQLAVHLKDLNAFDVELDAVHTNIVMVDVHADCTAAAVAADLT
AHGVRVSVFGPRRLRLVTHLDIDDAGVERTVAVFKRLYG

>72

MSTPRTTATAAKPKPYSEVNDYSVGMHPKILDLMARDNMTQHAGYQDQSHCAKAARLIGE
LLERPDAVHFISSGTQTNLIACSLALRPWEAVIATQLGHIISTHETGAI EATGHKVVVTP
CPDGKLRVADIESALHENRSEHMVI PKLVYISNTTEVGTQYTKQELEDISASCKEHGLYL
FLDGARLASALSSPVNDLTLADIARLDMFYIGATKAGGMFGEALII LNDALKPNARHLI
KQRGALMAKGWLLGIQFEVLMKDNLF FELGAHNSKMAAILKAGLEACGIRLAWPSASNQL
FPILENTMIAELNNDFMYTVEPLKDGTCIMRLCTSWATEEKECHRFEVLKRLVASTA

>73

MVEAELPPQYTTASNDLRSDTFTTPTKEMIAAAMNASIGDAVYGEDVDTVKLEQHVAKLA
GKEAGIFCVSGTSLNQIALRTHLYQPPYSILCDYRAHVYTHEAAGLAILSQAMVVPVIPS
NGDYMTLEDIKSHYIPDDGDIHGAPT KVISLENTLHGIVYPLEELVRIKAWSEMENGLKLH
CDGARIWNAAAETGVPLKQFGEIFDSISICLSKSMGAPMGSILVGNTKFIKKNHFRKQQ
GGGVRSQGMARMALCNVDDDWKSKLLYSHKLAHSLADFCEKNGIPLES PADTNFVFDL
RKARMDPDLVVKKGLKYNVKLMGGRVSFHYQVTEQTLEKVKLAILETFEYAKQHPFDQNG
PTQIYRSESTEIDIEGNAIKEIKTYKY

>74

MTDMELPPAYITASNDLRSDTFTTPTTEAMMKAALTASVGDVYGEDIDTVKLEQKVAKLA
GFEAGLFCVSGTLSNQIALRTHLYQPPYSVLCDYRAHVYTHEAAGLAILSQAMVVPVIPS
NGNYLTLQDIKRHYVPDDGDIHAAPT KVISLENTLHGIVYPLEELIRIKGWC MENGLK LH
CDGARIWNAAVQANIPLKQYGEIFDSISICLSKSIGAPMGSVLVGDVKFIKKNHFRKQQ
GGGVRQSGIMARMAMVAIEDNWKERLHRSHSLAHNLA AFCKKQGI PLESPADTNFVFDL
RKARMDPDLVKKGLKYNV KLMGGRVS FHYQIEDDVL ERVKLALSEAF EYAKQHPCDQDG
PTQVYRSASTEIDVHGNAIQEIKTYKY

>75

DFRSDVVTVPEEMMSTILEASVNDDIYDEPGDPSVKALEARLVELTGMEALWAVSGTQ
GNQICLRTHLTQPPHSVLLDHRRAHVHCWESGALPVISQASVTTVHAKNGVHLTLEDVKKN
IADGNIHFPPTRVSVLENTLSGTILPLADAQAISDYVRSFPVPEGQKPIAMHLDGARVF
DGVIGEGVDLKAYAACFDSISICLAKGIGAPMGSVILGKKSFIERAKWFRKMLGGGTRQP
GMMAAAALSALEYSIPRFPSVHAMTKDAAARLEAVGYMFTLPVQTNMILLDLEAVEIPPA
AFVEYCAKEDISVFPMARLVFHHQ TSEGAVDKLV TALTRLMEDKRNGVT LNDGEAHGGYS

>76

MIESSFQDDMYDATGDES VNRLQORLVELTGKEAMWALS GMTGNQICLRTHLTQPPHTV
LLDYRAHVHCWETGAMPVMSQA AVTQVHPKNGLHLTLEDVKKHIADGNIHFPPTRVSVL
ENSLSGTILPLKDAKEISDFVRNFPVEEGEKPIAMHLDGARLFD AVAAEGVSLKEYCACF
DSISICLAKGLGAPMGSIIVGSKKFIHRARLFRKMF GGTRQPGMMAAALCALEYTMPL
LPKVHALTKRTADSLREAGYKITLPVQTNMIVLDLEEVGIPPAAFVEY GK RAGVTVFPTG
RLVFHHQISAAAASRLVDAL TSLMQDKKAGRQLENH KVTGGYM

>77

MTADEEETFTTFNEFRSDTFTVPTRAMVEDGFMNATFGDSVYQEDETTLRLESKMCEITG
KPAALFCVSGTMSNQI GLRANLVQPPYSILCDYRAHVFLHEAGGLATLSQAMVHPVRPSN
GNYLTFEDVLGNVTYDDGDIHAAPT KVISLENTLHGIIPIIEEIRKISEFCRENDIRLH
LDGARLWNASVATGISIKEYCSYFDSVSLCLSKSLGAPIGSVLVGD EKFIRKANHFKKQS
GGGIRQAGIMSAMAHAIDYNLSKLELSHNYAKQIGDFCQEHGIKLES PVD TSLVFLDLK
ANKMDPNRLVELGR TKYNV KLMGQRIACHFQLSQESVDNVK KCI LECLEYHQKHPHKDDG
RNNKKMYSLDAIKK

>78

MTEFELPPKYITAANDLRSDTFTTPTAEMMEEAAL EASIGDAVYGEDVDTVRLEQTVARMA
GKEAGLFCVSGTLSNQIAIRTHLMQPPYSILCDYRAHVYTHEAAGLAILSQAMVVPVPS
NGDYLTLEDIKSHYVPDDGDIHGAPTRLISLENTLHGIVYPLEELVRIKAWCMENGLK LH
CDGARIWNAAAQSGVPLKQYGEIFDSISICLSKSMGAPIGSVLVGNLKFVKKATHFRKQQ
GGGIRQSGMMARMALVNINNDWKSQ LLYSHSLAHELAEYCEAKGIPLESPADTNFVFINL
KAARMDPDLVKKGLKYNV KLMGGRVS FHYQVTRDTLEKVKLAI SEAFDYAKEHPFDCNG
PTQIYRSESTEVDVDGNAIREIKTYKY

>79

MNQDMELPEAYTSASNDFRSDTFTTPTREMIEAALTATIGDAVYQEDIDTLKLEQHVAKL
AGMEAGMFCVSGTLSNQIALRTHLTQPPYSILCDYRAHVYTHEAAGLAILSQAMVTPVIP
SNGNYLTLEDIKKHYIPDDGDIHGAPTKVISLENTLHGIIHPLEELVRIKAWCMENDLRL
HCDGARIWNASAESGVPLKQYGEIFDSISICLSKSMGAPMG SILVGS HKFIKKNHFRKQ
QGGGVRQSGMMCKMAMVAIQGDWKGKMR RSHRMAHELARFCAEHGIPLESPADTNFVFLD
LQKSKMNPDLVKKSLKYGCKLMGGRVS FHYQISEESLEKIKQAILEAF EYSKKNPYDEN
GPTKIYRSESADAVGEIKTYKY

>80

MTEFELPPKYITAANDLRSDTFTTPTAEMMEEAAL EASIGDAVYGEDVDTVRLEQTVARMA
GKEAGLFCVSGTLSNQIAIRTHLMQPPYSILCDYRAHVYTHEAAGLAILSQAMVVPVPS
NGDYLTLEDIKSHYVPDDGDIHGAPTRLISLENTLHGIVYPLEELVRIKAWCMENGLK LH
CDGARIWNAAAQSGVPLKQYGEIFDSISICLSKSMGAPIGSVLVGNLKFVKKATHFRKQQ
GGGIRQSGMMARIALVNINNDWKSQ LLYSHSLAHELAEYCEAKGIPLESPADTNFVFINL

KAARMDPDLVKKGLKYNVVKLMGGRVSFHYQVTRDTLEKVKLAI SEAFDYAKEHPFDCNG
PTQIYRSESTEVDVDGNAIREIKTYKY

>81

DFRSDTVTRPTEQMLAAIAATTLQDDDFRQDPTTLGLEAWMAELTGKAAGLFFVVS GMTGN
QLGVR AHLQSPPHSVLCDARSHLVTHEAGGVASLSGAMVSCVTPVNGRYMTQADLEAHVN
RGTLLITDCPTRLVLEIPLGGVILPLDKCRRISEWARAQGIALHLDGARLWEAVAAGAGS
LRDYCACFDSVSLCFSKGLGAPIGSVLVGSSETLRERARWIRKSI GGGMRQAGVVCAARV
AVEATFLGGLLKRSHARARDIATFWEIHGGRLTYPTETNMVWLDLEAVGWTPERLIRRGA
ELGLRFMGARLVVHYQIGDEAIGRLQDLMLEILVS

>82

DFRTDTITMPTDEM FELMRDASRGDDVYGEDQATNDLQDKIAKMAGKEAGLFCVSGTMTN
QLAIRTHLTQPPHSIILDARSHVHLYEAGGVAMHSQASSHAVPPRNGHHLTLEEILDNAV
FGEDIHSAPT KLVSLENTLSGMVFPQEEIVRISDCMRNSNGIIMHCDGARMWEAAVKTGLS
LEELCRPFDTVSLCLSKGLGAPVGSVLVGPVKFIEKVKWFRKAFGGGIRQCGSLAVAADY
CVDNHFLKLGKTHELASYLAKSLADLGAQLLLPVE TNMLWLDPSSSLGFSIADLAARAKTR
GITLGSNRIVVHHQITREACEDLVALVAELKDE

>83

VNFLSDTVTCPTVGMQRVIAAAEVGDDVFGADPSVKRLEKVAERLGKPAALYVPSGTMS
NLIAIGTHCRRGDEVICGDKAHIFLYEGGASAYMGVSLHTVFNQPDGTLDIKDIHNAVR
DDDPHYPRTRLVEIENTQNTCGGRVPLSYIREVEQFCQERDLRLHVDGARLANASVASG
IPMDELARGADSVSLCLSKGLGAPVGSILAGSEEFIHHRRLRKS LGGMRQAGIIASAG
LYALENQFDRLEEDHVNAQALAHGISSIPGVEIDPSTVDTNIVFFTLTKDAKLDATTLVQ
KLGSEKGVLVGAYADGNRVRAVTNLHISAEDVEYTISSVRALLN

>84

VNFLSDTVTRPSAAMRQVIAAAEVGDDMFGADPSVKRLEKVAEQLGKAAALYVPSGTMS
NLIAIGTHCRRGDEVICGDKSHIFLYEGAGASAYMGVSLHTVFNQADGTLDIKDIHNAVR
DDDPHYPRTKLVEIENTQNTCGGRVPLSYIREVEQLCHERDLRLHVDGARLANASVASG
TPMDELVRGADSVSLCLSKGLGAPVGSVLGSEEFIHQAQRFRKSLGGMRQAGIIASAG
LYALENQFDRLVEDHVNAQALAHGISSIPGVETNLDTVDTNIVFFALTPDAKLNATTLVQ
KLESEKGVLVGAYADGNRVRAVTNLHISSQDVEYTVASIRALLS

>85

IDLRSDTVTRPDDRMREASANA EVGDDVYREDPTVNELEERAAEIVGTEAALYVPTGTMG
NQIAVRTHTERGQEVLCERESHVVKWELGGMAQLSGLQVRTVEGDDRGVVS PERVRESYV
AEDLHRP GTGLLTLENTHNSKGGAAIEVEKIRETAETAHGLDVPVHLDGARVFNAAAALG
VEAREIVAPVDSVMFCLSKGLGAPVGSMLAGSQEFVDRARRNRKLLGGMRQAGIIAAPG
ALALENRHRLTDDHENARHLAAGLDGIEGLSVLEPETNIVLVDC EGTGHTAEFFLAAVES
TGVLVGAFGEYVVRFC THWDVDDGDVEEAI SAVDRAL

>86

IDLRSDTVTTPGEAMRQSAADADVGDDVYGEDPTVNELEEARAADAVGMEAAALYVPSGTMG
NQIAARVHTERGQEVLEQESHVYKYELGGFAQHSQLQVRTYAGGANGCPTPEQVREGYV
EEELHRP GTGLLCLENTHNVKGGVPVPRDELAAAADAASNGLP VHVDGARVFNAATALD
VDAADLLAPADSVMFCLSKGLGAPVGSMLAGSEAFVERARRTRKLMGGMRQAGVIAAPG
LLALENRERLDVDHENARRLAAELDDLEGLSVPVPE TNIVLVDTTETGLSAAEFLENCEA
EGVLGSEFGEYTVR FCTHLDVDAADVDECVA AVERAR

>87

LDFRSDTVTNPTPEMREAAA KAPLGDDVYREDPSINRLEKISAEILGKEAGL FVTS GTMG
NAVAI LAHTQRGDEIILEEKSHIFMNEVGGLAVMGSLMARTISGD LGWMPEDIRAAVRA
ENIHYPKTSLVCIENTHNSAGGIALTVDMKMDWDVSKENELGVHLDGARVFNAAIALDV
NVKELTQYADTVQICLSKGLSAPVGSVVVGSYDLIEKARKYRKMLGGMRQAGIIAAPGI
IAITKMDRLAEDHENAKVLSEGLRNLGIKIVNVPQTNM VYIDLSSIGWTGKEFTEACAK
IGWKIRGASPIVRLCTHYGIEREDIDTFLEGMAKLVP

>88

IDLRSDTVTKPDDAMREAARTADVGDDVYGEDPTVNELEARVAEAVGKDAALYVPSGTMG
NQIAARVHTEGQEVLDRESHVVAYELGGFAQHANLQVRTLETDRGVPTPEQIAD EYVT
EDLHRPGTGLLCLENTHNARGGLAIDPERIAAAAEAAERHVPVHLDGARVFNAAMALDV
PVTEITDPEVDSVSVCLSKGLGAPVGSVLAGDEAFVERARRVRKLLGGGMRQAGI IAGPGL
EALENVSELESDHENAQALADGLAAVPGFDVREPETNIVLADISGTGLETA AVVERLRDR
DVLASEFGPATVRFCTHRDVS RADIGRALEKITETFE

>89

IDLRSDTVTLPSDEMREAAARDAAVGDDVYGGDDPTVNELEARAAELVGKEAALFVPSGTMG
NQIAARVHADPGQEALVDEKAHVYKWEVGGFAQLSGLQVRAYDAGERAAPTPAQVRDHAR
EESLHVAGTGVLCLENTHNARGGVAVPKADIDAAADAARDLGI PVHLDGARLFNACVALD
EDPTAMVERVDVTMCCLSKGLGAPVGSILAGPEAFIEEAVRVRKQFGGGMRQAGLIAAPG
LVALDNVDRLADDHANATVLAEGLDAVSGLSVPTPDTNIVVVDSEGAGLTAE EFVELCDD
VGVLGTFGRYHTRFTTNLNVSRADVERAVDLVADAVESR

>90

IDLRSDTVTLPSDEMREAAARDAAVGDDVYGGDDPTVNELEARAAELVGKEAALFVPSGTMG
NQIAARVHADPGQEALVDEKAHVYEWVGGFAQLSGLQVRAYDAGERAAPTPAQVRDHAR
EESLHVAGTGVLCLENTHNARGGVAVPKADIDAAADAARDLGI PVHLDGARLFNACVALD
EDPTAMVERVDVTMCCLSKGLGAPVGSMLAGPEAFIEEAVRVRKQFGGGMRQAGLIAAPG
LVALDNVDRLADDHANATVLAEGLDAVSGLSVPTPDTNIVVVDSEGAGLTAE EFVELCDD
VGVLGTFGRYHTRFTTNLNVSRADVERAVDLVADAVESR

>91

IDLRSDVTTTPSNAMRDAARDADVGDDVYGEDPTVNELEAAVADRLGFEEAAVFPVTGTMG
NQIAIRTHTEPGQEILVDREAHVFNWEVGGALHSGLQTRPFEE TVRRDRGRERGVPTPEA
IESGYVEESVHRAGTGVLSELENHNGRGLAIAPERIDAAAEAAHDLGVPVHLDGARLFN
AAVAHDVPAERFTRQVDSAMVCLSKGLGAPIGSMLAGSAEFVEAARRNRKLLGGGMRQAG
IVAAPGLLALDSVDRLADDHRNADRLAVGLADLPGLDVVAPETNIVLVDVSDAAQDAAAF
KQACADAGVACTTMDETTARFCTHRNVDKADVDEALS LVADVVD

>92

IDLRSDVTTTPSEAMREAAARDAEVGDDVYRDDPTVNELEERRAAEAVGTEAALYVPSGTMG
NQIAVHVHTEPGQELILERESHYRWELAGAAKLSGTQTRTL DAGERCVPTPEAVREGLV
DEDLHRPGTGLLALENTHNRYGGTAIPVDRI SAAAEVARDAGVPVHLDGARVFNA AVALG
VDASEIVDPVDSVTFCLSKGLGAPVGSILAGDEAFVEAARRVRKLFGGGMRQAGMIAAPG
LLALENVDRDLADDHANAERLAAGLDALDGVRAPEPDTNIVVAHTEDAGIAAADLV TACKD
AGVGCVEFDDYATRFTTHLDVDGDDVDAAIDRIGDVVAELA

>93

MSNNKGFASDNHSGIHPDILKAIISANVGHANAYGNDDYTKRAVKKFEEYFGSNIEVYFV
YNGTAA NILGLKTLTDSFNSIICAETAHLNVHECCGPENFIGCKLITIP TVDGKLTINQI
KPHIFGFEDPHRAQPKIISITQPTELGTVYTP EEEIRKLSNFAHRNRMLIHIDGARLCNAA
AYLDVEMKDI TGNVGDILSFGGTKNGMMFGEAVIFFNKQSSKNFV FIRKQGMQLPSKMR
FISSQFETLLSNDLWLNKAKHSNRMAQLLYKEIRDVPQIKITQKVETNAIFAIVPKKYIH
LLQKEYFFHVNNERISEVRWMC SFDTTKKDIMNFSEI I KKTVV

>94

VDLRSDVTRPTDAMREAMCNAEVDDDVLGYDPTARRLEEEEMAKMMGKEAALFVPSGTMG
NLISVMVHCDVRGSEVILGDNCHIHVYENG GISTIGVHPKTVKNEEDGTM DLEAIEAAI
RDPKGSTFYFSTRICLENTHANS GGRCLSVEYTEKVGEIAKRHG VKLHIDGARLFNASI
ALGPVHKLKVAADS VQVCLSKGLGAPVGSVIVGSQSFIEKAKTVRKT LGGGMRQIGVLC
AAALVALQENLPKLQHDHKKAKLLAEGLNQMKGIRVNVA AVETNMIFMDMEDGSRFTA EK
LRKNLEENGILLIRGNSSRIRIVIHQITTS DVHYTLSCFQQAMLMQ

>95

VDLRSDVTKPTESMRSAMANA EVDDDVLGNDPTALRLEKEVAE IAGKEAAMFVPSGTMG
NLISVLVHCDER GSEVILGDDSHIHIYENGGVSSLGGVHPRTV KNEEDGTMEIGAIEAAV

RSPKGLDHPVTKLICLENTQANCGGRCLPIEYIDKVGELAKKHGLKLHIDGARIFNASV
ALGVPVKRIVQAADSVSICLSKIGIGAPVGSVIVGSKKFI TKARWLRKTLGGGMQRQIGLLC
AAALVALHENVAKLEDDHKKARVLAEGLNRIERLRVNVA AVETNIIYVDI PEDPKFGAEE
ACKSLEDVGVLVIPQATFRIRIVLHHQISDQVDEYVLSCFEKIFHS

>96

VDLRSDTVTRPGPAMRRAMAEAVVGDDDYGEDPTVHELQEKAAELLGVERTLFVPTNTMA
NLISVMGHCRRRGSQVLLGQECHLHVYEQGGVAQIAGVHSHPLPDLPHYGTLDLNELEAL
TRGSGSPYHPVCELVYLENTHSSAGGRVLPVDYLRQVCLLAHAHGARVHMDGARLMNAAV
ALRIPPARLVEHCDSVSFCFSKGLGAPV GALVGGSKDFIGEAWRLRKALGGGMQRQAGVLA
AAALVGLAEAEVLP RDHENAQRFAKGLQDLASPICSVDPATVETNMVLVQVAGLPPSEL
CQRLQAVSAEEVAQTGRAVRVLLFPWTEQSVRAVWHRDVS AQDTEALAKKWEFVLRQLR

>97

MNGETSRPPALGFSSDNIAGASPEVAQALVKHSSGQAGPYGTDELTAQVKKRFCEIFERD
VEVFLVPTGTAANALCLSAMTPPWGNIYCHPASHINNDECGAPEFFSNGAKLMTVDGPAA
KLDIVRLRERTREKVG DVHTTQPACVSI TQATEVGS IYTLDEIEAIGDVCKSSSLGLHMD
GSRFANALVSLGCS PAEMTWKAGVDALSFGATKNGVLAEEAIVLFNTSLATEMSYRRKRA
GHLSSKMRFLSAQIDAYLTDDLWLRNARKANAAAQRLAQGLEGLGGVEVLGGTEANILFC
RLDSAMIDALLKAGFGFYHDRWGPVVRVFTSFATTAEDVDHLLNQVRLAADR

>98

LA AIDELVAEEEEADARVLHLTANETVLS PRARAVLASPLTSRYLLEHLDMRGPSPARLG
NLLLRLGLDRIGTIEESATEVCRRLF GARYAEFRCLSGLHAMQTTFAALSRPGD TVMRVAT
KDGGHFLTTELICRSFGRRSCTYVFDDTMTIDLERTREVV EKERPSLLFVDAMNYLFPFPI
AELKAIAGDVPLVFDASHTLGLIAGGRFQDPLREGADLLQANTHKTFFGPQKGIILGNDR
SLMEELGYTLSTGMVSSQHTASTVALLIALHEMWDGREYAAQVIDNARRLAGALRDRGV
PVVAEERGF TANHMFVDRPLGSGPAVIQRLVRAGVSANRAVAFNHLDTIRFGVQEITR
RGYDHDDLDEAADLVA AVLLERQEPERIR

>99

MGASGKTS LGVDPLMLGRAIIDADRRAAHALNLVPS ENRISPLASPLGSDFYNRYFFNT
DGDPLFWEFRGGEEIAHIEALGIDALRRMASARYCNVRPISGMSAMIITVAALSRPGSTV
VSVDQNSGGHYATSALLGRFGRESRLGGGGGRVDESRLADLLAPGDVLDVYVDVQNCVR
TPDFRAMSAVVKDVSPGTRLYVDASHYLG LVFGGHV VNP LTCGADAFGGSTHKSFPGPHK
GVIFNTSDDVDEKLRAAQFDLLSSHFAETLALALAALEVEQHI GEYARATNDNARRLAR
ALADAGFRVHGDSSAGYTDTHQVWVELDNTADAYALS NR LADVGIRVNLHSTLPGVPGVH
LRLGSNEVT FEGAGPRAIERLADALVTARE RALEPRTVSEIRRRFGAPFYTDSEK

>100

VDRLKSWLSL TEESREALNLVPS ENRMSPLAMLPLSSDFY NRYFFNDRLDSGFWQFRGG
QSASHFEVDVALPSLRRLTASEYVNLRPISGLNTMLIAVAGLGGVPGSAVVSIGHDWGGH
YATSALVERLGLRSVTVRVERGKVDASELERVLRQDAPTLVYLDLQNSLDPLDVPVAAA
IRTHSPGTL LHVDASHVLGLILGGAI PNPLDAGADSVGGSTHKTFFPGPHKGILFTRDAEI
ADRLRRAQFNLLSSHFAETISLGLAALEFENFGAAYAEQVLRNATAFGRALRHRGFDVI
GDPERTGTHQVWVRIGDAPRTDHAENLYLAGIRVNVQTDLPGLPGAMFRLGVSEVTFE
GADEDAMGLLADAFVHAADGRPGRAAGIRADIRAGMTRPYFYTGAE

>101

MTDTNELRKVLHRFRSQQEKA EFSVNLVPS ENKLSPLAQPLRSDY NRYFFNDALDPGF
WQFRGGQDVAEMETELTVDHLSRLARAPHVNERPISGLSAMIAMAGLGGKPGGT VVSID
AASGGHYATASMARLGFESATVPVVRGQVDEQRLEQVLCRHEPELIYLDLQNSRHELEV
SRVADLVGRHSRQTL LHVDCSHTMGLVLGGALGNPLDAGANTMGGSTHKTFFPGPHKGVLF
TRTPELHQRMREAQFTMLSSHFAETLALGLASAEFSHF GPAYAEQVIGNAQLFSKLLAS
EGFDVATDEDGHTTSTHQIWKIGDAEQTDRI SQSLYDHGIRVNVQVDLPGMPPGVLRLLG
ISELTFVGGREAAVHALAREFSNARAGVRRD GSGSRRVREQCGSPFFHFVDYP

>102

MTDIRELRKVVDRFRAQERKAAASINLVPS ENKLSPLAQMPLSTDY NRYFFNDELDPGF

WQFRGGQEVAKIQTE LARGHL SRLARAPYVNERPISGLSAMMMAMAGLGGPPGGTVVSI D
AASGGHYATADMARRLGFESATVPVVRGRVDEQWFGQVLRHVPELVYLDLQNSRHELEV
SRVAELIEAHSPTILHVDCSHTMGLILGGALSNPLDAGAHTMGGSTHKSFPGPHKGVLF
TRSPELHQRLKHAQFTMLSSHFAETLALGLAAAEFRHFHGHAYAEQV VANARLLGKLLAA
DGFDVTADENGHATSTHQWLWVRIGDAEQTRDRFSKYLYDHGIRVNVQVDLPGLPGPVLRLG
VNELTFLGGHEAAVHALAEFFSHARDGVRRDGEGSQRVREQYGPFFYFVEFS

>103

MTGIKELRDVDRFRAEERKAATAVNLPSENRLSPLAQLPLSTDYNNRYFFNDALDPGF
WQFRGGQEVAEIQTE LARGHL SRLSRAPHVNERPISGLSAMMMALAGLGGKPGGTVVSVG
AESGGHYATAGMARRLGFESATVPVAHGQVDEQRLGQLLRERTPQLLYLDLQNSRHELEV
SRVAELIKEYSPSTLLHVDCSHTMGLILGSALGNPLDAGADTMGGSTHKTFFGPHKGVLF
TRSPELHQRLKDAQFTMLSSHFAETLSLGLAAAEFHHFQAYAEQVIANARLFSKLLAA
DGFDVAADENGHATSTHQVWVKIGDAERTDRISQALYEHGIRVNVQVDLPGLPGPALRLG
VNELTFTGGREAAVHALAEFFGNARAGVRRDGDGARRVCEQSGPPFFYFAEFS

>104

MFSRDLTIKDYDADLFAAMEQEAVRQEEHIELIASENYTS PAVMEAQGSVLTNKYAEGYP
GKRYYGGCEYVDVVEQLAIDRAKELFGADYANVQPHAGSQANSAYLALLQGGDTILGMS
LAHGGHLTHGASVSSGKLYNAVQY GIDANGLIDYDEVERLAVEHKPKMIVAGFSAYSQI
LDFPRFRAIADKVGAYLFVDMAHVAGLVAAGVY PNPVYADVVT TTTTHKTLRGPRGGLI L
ARANAIEKKLNSAVFPGAQGGPLEHVIAAKAICFKEALQPEFKTYQQQVVKNAKAMAGV
FIERGFDVVSGGTENHLFLLSLIKQDISGKDADAALGKAFITVNKNSVPNDPRS PFVTS G
LRFGTPAVTTTRGFKEAECKELAGWICDILADLNNEAVIDAVREKVKAI CKQLPVYGA

>105

MFSKQDQIQGYDDALLAAMNAEEQRQEDHIELIASENYTSKRVMQAQGSGLTNKYAEGYP
GKRYYGGCEHVDKVEALAIERAKQLFGADYANVQPHSGSSANSAYLALLINPGDTILGMS
LAHGGHLTHGAKVSSGKLYNAVQY GINTDTGLIDYDEVERLAVEHQPKMVVAGFSAYS K
TLDFPRFRAIADKVGALLFVDMAHVAGLVAAGLYPNPLPYADVVT TTTTHKTLRGPRGGLI L
LAKANEAEIKKLNAAVFPGAQGGPLMHVIAGKAVCFKEAQEPGFVKYQQQVIDNAQAMAE
VFIKRGYDVVSGGTDNHLFLVSLIRQGLTGKDADAALGRAHITVNKNAVNDPQSPFVTS
GLRIGTPAVTTTRGFKVAQCIELAGWICDILDNLGDADVEANVAKHVSALCADFPVYR

>106

MFSKHDQIRGYDDELLAAMDAEEARQEDHIELIASENYTSKRVMQAQGSGLTNKYAEGYP
GKRYYGGCEHVDKVERLAIDRARQLFGADYANVQPHSGSSANAAYLALLNAGDTILGMS
LAHGGHLTHGAKVSSGKLYNAVQYGLDTATGLIDYDEVERLAVEHKPKMIVAGFSAYS K
TLDFPRFRAIADKVGALLFVDMAHVAGLVAAGLYPNPIPFADVVT TTTTHKTLRGPRGGLI L
LARANEEIEKKLNSAVFPGAQGGPLMHVIAAKAVCFKEALEPGFKDYQAQVIRNAKAMAE
VFIGRGYDVVSGGTDNHLMLISLVRQGLTGKEADAALGRVGITVNKNAVNDPQSPFVTS
GIRIGTPAVTTTRGLQEAQSRELAGWICDILDHLGDADVEAKVATQVAGLCADFPVYR

>107

MFSKHDQLQGYDDELLAAMDAEDRRQEDHIELIASENYASKRVMQAQGGGLTNKYAEGYP
GKRYYGGCEHVDKVERLAIDRARQLFGADYANVQPHSGSSANAAYLALLNAGDTILGMS
LAHGGHLTHGAKVSSGKLYNAVQYGLDTATGLIDYDEVERLAVEHKPKMIVAGFSAYS K
TLDFPRFRAIADKVGALLFVDMAHVAGLVAAGLYPNPIPFADVVT TTTTHKTLRGPRGGLI L
LARANEEIEKKLNSAVFPGAQGGPLMHVIAAKAVCFKEALEPGFKDYQAQVIRNAKAMAE
VFIGRGYDVVSGGTDNHLMLISLVKQGLTGKAADAALGAAHITVNKNAVNDPQSPFVTS
GIRIGTPAVTTTRGFREGECRELAGWICDILDDIDNPEVGERVVRGQVGEFCRHFVYAD

>108

MFSRDLTLARYDAELFAAMEQEQRQEEHIELIASENYTS PAVMEAQGSVLTNKYAEGYP
HKRYYGGCEYVDIVEQLAIDRAKQLFGADYANVQPHAGSQANAAYLALLSAGDTILGMS
LAHGGHLTHGASVSSGKLYNAVQY GIDANGLIDYDEVERLAVEHKPKMIVAGFSAYSQV
LDFARFRAIADKVGAYLFVDMAHVAGLVAAGVY PNPVFPADVVT TTTTHKTLRGPRGGLI L
ARANEEIEKKLNSAVFPSAQGGPLEHVIAAKAVCFKEALQPEFKTYQQQVLKNAQSMQV
FLDRGFDVVSGGTQNHFLFLLSLIKQDITGKDADAALGRAFITVNKNSVPNDPRS PFVTS G

LRIGTPAVTTRGFKEAECRELAGWICDILENMGDES VVDGVREKVKAI CAKFPVYGN

>109

MFSKKDQIQGYDDALLAAMNAEEQRQEDHIELIASENYTSKRVMQAQGSGLTNKYAEGYP
GKRYYGCEHVDKVEALAIERAKQLFGADYANVQPHSGSSANSVYLLALLNAGDTILGMS
LAHGGHLTHGAKVSSSGKLYNAVQYGIN TDTGLIDYDEVERLAVEHKPKMIVAGFSAYSK
TLDFPRFRAIADKVGALLFVDMAHVAGLVAAGLYPNPLPYADVVT TTTTHKTLRGPRGGLI
LAKSNEEIEKKLNAAVFPGAQGGPLMHVIAKAVCFKEAQEPGFKVYQQQVIDNAQAMAS
VFIKRGYDVVSGGTDNHLFLVSLIRQGLTGKEADAALGRAHITVNKNAVNDPQSPFVTS
GLRIGTPAVTTRGFKVPQCIELAGWICDILDNLGDADVEANVAKHVSALCADFPVYR

>110

MFSRDLTIAKYDADLFAAMEQEAVRQEEHIELIASENYTS PAVMEAQGSVLTNKYAEGYP
GKRYYGCEYVDVVEQLAIDRAKELFGADYANVQPHAGSQANSVYLLALLQGGDTILGMS
LAHGGHLTHGASVSSSGKLYNAVQY GIDANGLIDYDEVERLAVEHKPKMIVAGFSAYSQI
LDFPRFRAIADKVGAYLFVDMAHVAGLVAAGVYPNPVYADVVT TTTTHKTLRGPRGGLIL
ARANAIEKKLN SAVFPGAQGGPLEHVIAAKAICFKEALQPEFKTYQQQVVKNAQTMASV
FIERGFDVVSGGTENHLFLSLIKQDISGKDADAALGKAFITVNKNSVNDPRS PFVTS
LRFGTAVTTRGFKEAECKELAGWICDILADLNNEAVIDAVREKVKAI CKKLPVYGA

>111

MFSRDLTIAKYDAELFEAMQOEALRQEEHIELIASENYTS PAVMEAQGSVLTNKYAEGYP
GKRYYGCEYVDVVEQLAIDRAKELFGADYANVQPHAGSQANAAYLLALLSAGDTILGMS
LAHGGHLTHGASVSSSGKLYNAIQY GIDGNGLIDYDEVERLAVEHKPKMIVAGFSAYSQV
LDFARFRAIADKVGAYLFVDMAHVAGLVAAGVYPNPVYADVVT TTTTHKTLRGPRGGLIL
ARANADIEKKLN SAVFPGAQGGPLEHVIAAKAICFKEALQPEFKAYQQQVVKNAQAMASV
FIERGFDVVSGGTQNHFLFLSLIKQDISGKDADAALGKAFITVNKNSVNDPRS PFVTS
LRFGTAVTTRGFKEAECKELAGWICDILADLNNEAVIDAVREKVKAI CKKLPVYGN

>112

MFSKQDQIQGYDDALLAAMNAEEQRQEDHIELIASENYTSKRVMQAQGSGLTNKYAEGYP
GKRYYGCEHVDKVEALAIERAKQLFGADYANVQPHSGSSANGAVYLLALLQAGDTILGMS
LAHGGHLTHGAKVSSSGKLYNAVQY GIDTNTGLIDYDEVERLAVEHKPKMIVAGFSAYSK
TLDFPRFRAIADKVGALLFVDMAHVAGLVAAGLYPNPIPFADVVT TTTTHKTLRGPRGGLI
LAKSNEEIEKKLNAAVFPGAQGGPLMHVIAKAVCFKEALEPGFKAYQQQVIENAQAMAQ
VFIDRGYDVVSGGTDNHLFLVSLIRQGLTGKDADAALGRAHITVNKNAVNDPQSPFVTS
GLRIGTPAVTTRGFKVAQCVALAGWICDILDNLGDADVEADVAKNVAALCADFPVYR

>113

MFSRDLTIAKYDADLFAAMEQEQRQEEHIELIASENYTS PAVMEAQGSVLTNKYAEGYP
GKRYYGCEYVDVVEQLAIDRAKELFGADYANVQPHAGSQANAAYLLALLQGGDTILGMS
LAHGGHLTHGA AVSSSGKLYNAIQY GIDANGLIDYDEVERLAVEHKPKMIVAGFSAYSQI
LDFPRFRQIADKVGAYLFVDMAHVAGLVAAGVYPNPVYADVVT TTTTHKTLRGPRGGLIL
ARANADIEKKLN SAVFPGAQGGPLEHVIAAKAICFKEALQPEFKAYQE QVVKNAQAMAEV
FIARGFDVVSGGTKNHLFLSLIKQDISGKDADAALGKAFITVNKNSVNDPRS PFVTS
LRFGTAVTTRGFKEAECKELAGWICDILADLNNEAVIDAVREKVKAI CKKLPVYGA

>114

MFSKQDQIQGYDDALLAAINAEQRQEDHIELIASENYTSKRVMQAQGSGLTNKYAEGYP
GKRYYGCEHVDKVEALAIERAKQLFGADYANVQPHSGSSANSEVYLLALLQAGDTILGMS
LAHGGHLTHGAKVSSSGKLYNAVQY GIDTRTGLIDYDEVERLAVEHKPKMIVAGFSAYSK
TLDFPRFRQIADKVGALLFVDMAHVAGLVAAGLYPNPLPYADVVT TTTTHKTLRGPRGGLI
LAKANEIEKKLNAAVFPGAQGGPLMHVIAKAVCFKEALEPGFKAYQQQVIDNAQAMAG
VFIKRGYDVVSGGTDNHLFLVSLIRQGLTGKDADAALGRAHITVNKNAVNDPQSPFVTS
GLRIGTPAVTTRGFKVTCQTELAGWICDILDHLGDADVEANVARQVAALCADFPVYR

>115

MAGKRPFAGARFHDEIERTAALIACRVFNAE HANLQPHSCSQANQSVYHALLEPGDNVLA
LNFKAGGHLTHGHKVNFSGMFFNFRHYGVDEATDLIDYDLAEQDAIRFKPKLIVCGSSSY

PRLFDARRLREISDKVGALLMFDLSHEAGLIACGAI PNPVPLADVATMSMDKTMRGHGA
I IILCTAKIAQKIDKGVHPGTQSSFPISRLTQTAQALLHSQTAEFREYANRVLDNALLLEQ
HFLCIPNLLVTTGGTDKHYLVLNTKAAFIDGVLAEQRLEAISVLSRQTLPGDRTSRIDD
AGGIRLGTAWITSRGYELDEVSAIATIIIEALSPSFDDAKKHHLLSRVNTLIATDKPKDV
WRNS

>116

MFSRDLTI AKYDADLFAAMEQEALRQEEHIELIASENYTS PAVMEAQGSVLTNKYAEGYP
GKRYYGCEYVDVVEQLAIDRAKELFGADYANVQPHAGSQANSVYLLALLQGGDTILGMS
LAHGGHLTHGASVSSSGKLYNAVQY GIDGNGMIDYDEVERLAVEHKPKMIVAGFSAYSQI
LDFPRFRAIADKVGAYLFVDMAHVAGLVAAGVYPNPV PFADVVT TTTTHKTLRGPRGGLIL
ARANADIEKKLNSAVFPGSQGGPLEHVIAAKAICFKEALQPEFKTYQQQVVKNAKAMAGV
FIERGFVSVGGTENHLFLSLIKQDISGKDADAALGRAFITVNKNSVPNDPRSPFVTS
LRFGTPAVTTTRGFKETECKELAGWICDILADLNNEAVIDAVREKVKAI CAKLPVYGA

>117

MFSKQDQIQGYDDALLSAMNAEEQRQEDHIELIASENYTSKRVMQAQGSGLTNKYAEGYP
GKRYYGCEHVDKVEQLAIERARQLFGADYANVQPHSGSQANAAVYLLALLQAGD TVLGMS
LAHGGHLTHGAKVSFSGKLYNAVQY GIDTTGLIDYDEVERIAVECCPKMLIAGFSAYSK
TLDFPRFRAIADKVGAYLFVDMAHVAGLVAAGLYPNLPYADVVT TTTTHKTLRGPRGGLI
LAKANEELKFKNSAVFPGGQGGPLMHVIAAKAVCFKEAMEPGFKTYQQQVIDNAQAMAQ
VFITRGTDFVSVGGTDNHLFLVSLIRQGLTGKEADAALGRAHITVNKNSVPNDPQSPFVTS
GLRIGTPAVTTTRGFKVTQCIELAGWICDILDNLGAADVEANVASQVAALCADFPVYR

>118

MLKREMNIADYDAELWQAMEQEKVRQEEHIELIASENYTS PRVMQAQGSGLTNKYAEGYP
GKRYYGCEYVDIVEQLAIDRAKELFGADYANVQPHSGSQANFAVY TALLEPGD TVLGMN
LAHGGHLTHGSPVNFSGKLYNIVPYGIDATGHIDYADLEKQAKEHKPKMIIGGFSAYS
VDWAKMREIADSIGAYLFVDMAHVAGLVAAGVYPNPV PHAHVVT TTTTHKTLGPRGGLIL
AKGGSEELYKKNLSAVFPGGQGGPLMHVIAKAVALKAEPEFKTYQQQVAKNAKAMVE
VFLERGYKVVSGGTDNHLFLVDLVDKNLTGKEADAALGRANITVNKNSVPNDPKSPFVTS
GIRVGTPAITRRGFKAEAEKELAGWMCVLD SINDEAVIERIKGKVLDCARYPVYA

>119

MFSKQDQIQGYDDALLAAMNAEEQRQEDHIELIASENYTSKRVMQAQGSGLTNKYAEGYP
GKRYYGCEHVDKVEALAIERAKQLFGADYANVQPHSGSSANA VYLLALLQAGD TVLGMS
LAHGGHLTHGAKVSSSGKLYNAVQY GIDTNTGLIDYDEVERLAVEHKPKMIVAGFSAYSK
TLDFPRFRAIADKVGALLFVDMAHVAGLVAAGLYPNP I PFADVVT TTTTHKTLRGPRGGLI
LAKSNEEIEKKLNAAVFPGAQGGPLMHVIAAKAVCFKEALEPEFKAYQQQVIENAQAMAQ
VFIDRGYDVVSGGTDNHLFLVSLIRQGLTGKDADAALGRAHITVNKNV PNDPQSPFVTS
GLRIGTPAVTTTRGFKVAQCVALAGWICDILDNLGDADVEADVAKNVAALCADFPVYR

>120

MFNDRMNIADYDPELWQSITDEVQRQEDHIELIASENYTS PRVMEAQGSGLTNKYAEGYP
GKRYYGCEYVDVAESLAIERAKSLFGADYANVQPHSGSQANAAVY QALCAPGD TVLGMS
LAHGGHLTHGSHVSFSGKMYNAVQY GITPETGILDYAEIERLAVEHKPTMI IAGFSAYS
IVDWAKFREIADKVGAYLFVDMAHVAGLVAAGLYPNP VPFADVVT TTTTHKTLGGPRGGLI
LAKANEAEIEKKLNSAVFPGQGGPLMHVIAAKAVAFKECAEPEFAVYQQQVLDNAKAMVK
SFLARGYKIVSGGTENHLFLVDLIAQDITGKEADAALGNAHITVNKNSVPNDPRSPFVTS
GLRIGTPALARRGVNAQQSAELALWMCVLDIAIKDEAKLATTITAVKVKVAALCKACPVY
G

>121

FFSAHLAETDPEIAKAISQELGRQQHEIELIASENIVSRVLEAQGSVLTNKYAEGYPGR
RYYGCCQFVDIAEELAIDRAKRLFGCGFANVQPNSGSQANQGVFMALMQPGD TFLGLDLA
AGGHLTHGAPPNVSGKWFKPVSYTVRREDQRIDMEQVERLAQEHKPKVI IAGGSGYPRHW
DFAKFREIADSVGAYFFVDMAHFAGLVAAGLHSPFP PHAHVAT TTTTHKTLRGPRGGMILT
NDEALAKKFNSAIFPGLQGGPLMHVIAAKAAAFGEALKPEFKIYAKQVIDNARALADTII
SGGYDITSGGTDNHLMLVDLQKKGLTGKAAEAALS RADITCNKNGVFPDFQKPTITSGIR

LGTPASTTRGFGVAEFKQVGS LIVQVLDGIAEKDGGDAAVEAAVKEKVHALTDRFPPIYA

>122

FFSQSLAERDASVRGAILKELERQSQVELIASENIVSRAVLDAQGSVLTNKYAEGYPGK
RYYGGCEFADEVEALAIERVKRLFNAGHANVQPHSGAQANGAVMLALAKPGDTVLGMSLD
AGGHLTHGAKPALSGKWFNALQYGVSRDTMLIDYDQVEALAQHKKPSLI IAGFSAYPRKL
DFARFRAIADSVGAKLMVMDMAHIAGVIAAGRHANPVEHAHVVTSTTHKTLRGPRGGFVLT
NDEEIAKKINS AVFPGLQGGPLMHVIAGKAVAFGEALTDDFKTYIDRVLANAQALGDVVK
AGGVLDVLTGGTDNHLVLDLRPKGLKGAQVEQALERAGITCNKNGIPFDPEKPTITSGIR
LGTPAGTTRGFGAAEFREVGRLLILEVFEALRTNPEGDHATEQRVREIFALCERFPPIY

>123

MLKREMNIADYDAELWQAMEQEKVRQEEHIELIASENYTS PRVMQAQGSQLTNKYAEGYP
GKRYGGCEYVDVVEQLAIDRAKELFGADYANVQPHSGSQANFAVYTALLQPGDTVLMGN
LAQGGHLTHGSPVNFSGKLYNIVPYGIDESGKIDYDEMAKLAKEHKPKMI IGGFSAYS
VDWAKMREIADSIGAYLFVDMAHVAGLIAAGVYPNPVPHAHVVTSTTHKTLGPRGGLIL
AKGGDEELYKKNLSAVFPQAQGGPLMHVIAGKAVALKEAMEPEFKVYQQQVAKNAKAMVE
VFLNRGYKVVSGGTENHLFLDLVDKNTLGKEADAALGRANITVNKNSVPNDPKSPFVTS
GIRIGSPAVTRRGFKAEVKELAGWMCVLDNINDEATIERVKAKVLDICARFPVYA

>124

SNAMSLMFMDKEIFDLTNKELERQCEGLEMIASENFTLPEVMEVMGSI LTNKYAEGYPGK
RYYGGCEFVDEIETLAIERCKKLFNCKFANVQPNSSGQANQGVYAALINPGDKILGMDLS
HGGHLTHGAKVSSSGKMYESC FYGVELDGRIDYKREIAKKEKPKLIVCGASAYARVID
FAKFREIAD EIGAYLFADIAHIAAGLVVAGEHSPFPYAHVVSSTTHKTLRGPRGGI IMTN
DEELAKKINS AIFPGIQGGPLMHVIAAKAVGFKFNLSDEWKVYAKQVRTNAQVLANVMD
RKFKLVSDGTDNHLVMSFLDREFSGKDADLALGNAGITANKNTVPGEIRSPFITSGRLR
GTPALTARGFKEKEMEIVSNYIADILDDVNNEKLQENIKQELKKLASNFI IYERAMF

>125

IFNNNLHETDKEINEI IKHEKLRQSSVIELIASENFVSPAVLEAQGALLTNKYAEGYPSK
RFYNGCEEVDKAENLAIERVKKLFNCKYANVQPHSGSQANQAVYLALLQPGDTVLMGSLD
SGGHLTHGAAPNMSGKWFNAVSVNKETYLI DYDEIERLADLHKPKLLIAGFSAYPRNI
DFAKFREIVDKVGAYFMADIAHIAAGLVATGEHQSPIPYAHAVTSTTHXTLRGPRGGLILS
NDEEIGHKINSALFPGLQGGPLMHVIAAKAVAFLENLQPEYKSYIQQVINSAKALASSLQ
ERGYDILTGGTDNHLVLDLRKDGITGKLAANSLDRAGITCNKNAIPFDETS PFITSGIR
LGTPACTTRGFKEKDFVLVGHMVADILDGLKNNEDNSALEQQVLNEVTKLIELFPFYG

>126

MFDRAQSTIANVDPEIFAAIEQENRRQEDHIELIASENYTS PAVMAAQGSQLTNKYAEGY
PGKRYGGCEYVDVVEQLAIDRVKQLFGAEAAVQPNSSGQANQGVFFAMLKPGDTIMGM
SLAHGGHLTHGSPVNMMSGKWFNVVSYGLNENEDIDYDAAEKLANEHKPKLIVAGASAFAL
KIDFERLAKIAKSVGAYLMVMDMAHYAGLIAAGVYPNPVPHADFVTTTTHKSLRGPRGGVI
LMKAEYEKPINSAIFPGIQGGPLMHVIAAKAVAFKEALSPEFKYQQKVVENARVLAETL
VKRGLRIVSGRTEHVMLVLDLRKHIITGKAAEAALGAAHI TVNKNAIPNDEKPFVTS
RLGSPAMTTRGFGPAEAEQVGNLIADVLENPEDAATIERVRAQVAELTKRFPVYR

>127

MLKREMNIADYDAELWQAMEQEKVRQEEHIELIASENYTS PRVMQAQGSQLTNKYAEGYP
GKRYGGCEYVDIVEQLAIDRAKELFGADYANVQPHSGSQANFAVYTALLEPGDTVLMGN
LAHGGHLTHGSPVNFSGKLYNIVPYGIDATGHIDYADLEKQAKEHKPKMI IGGFSAYS
VDWAKMREIADSIGAYLFVDMAHVAGLVAAGVYPNPVPHAHVVTSTTHKTLGPRGGLIL
AKGGSEELYKKNLSAVFPGGQGGPLMHVIAGKAVALKEAMEPEFKTYQQQVAKNAKAMVE
VFLERGYKVVSGGTDNHLFLVDLVDKNTLGKEADAALGRANITVNKNSVPNDPKSPFVTS
GIRVGTPAITRRGFKEAEAKELAGWMCVLDNINDEAVIERIKGVLDICARYPVYA

>128

MFSRDVRL ETYDPELAKAIAAEAGRQEDHVIELIASENYCSPLVMEAQGSQLTNKYAEGYP
GKRYGGCEFVDIAEQLAIDRIKQVFGADYANVQPHSGSQANQAVYLALLQPGDTILGMS

LAHGGHLTHGAKVNVSGKLFYAVQYGVNEQGLIDYDEVQRLATEHKPKMVVAGFSAYSQK
IDWARFRAIADSVGAYLFVDMAHIAAGLVAAAGVYPSMEHAHVVTSTTHKTLRGPRGGIIV
AKGASEELQKQLQSIVFPGIQGGPLMHVIAAKAVAFKEALEPAFKTYQQQVVKNAQAMAN
TLIARGYKIVSGGTENHMLVDMIGRVDVSGKDAEALGKAHITVNKNSVPNDPRSFPVTS
GLRLGTPAITTRGYQEQDSIDLANWIADVLDAPTDEAVLAKVRDAVTAQCKRYPVYG

>129

YAEKSPALWDAIRQEEKRQNTIELIASENIVSDAVREAQGSVLTNKYAEGYPGRRYYGG
CQYIDQVEQLAIDYAKKLFNAKFANVQPHSGSQANMAVYQALLKPGDVIILGMMDAGGHL
THGAKVNFSGKEYKSYEYGLNVETEELDFDQIRKVALEVKPKLIVAGASAYSRIIDWQKF
RDIADVEGAYLMVDMAHIAAGLVATDQHPSPIPVADIVTTTTHKTLRGPRGGMILSNLEI
GKKINSALFPGIQGGPLEHVIAGKAQAFYEDLQPQFTDYIKQVVKNAKAMAEEVFDESENI
RVVSGGTDNHLMIIDI TDTGLTGKDAQNLLDFVNITTNKESIPGDKRSPFITSGLRIGTP
AITSRFGNEEDARKTASLIEILSDPDNEATIEHVKKEVHELTKKHPVE

>130

MDVLAALERKPSLNLFPPIENRLSPRASAALATDAVNRYPYSETPVAVYGDVTGLAEVYAY
CEDLAKRFFGARHAGVQFLSGLHTMHTVLTALTTPGGGRVVLAPEDGGHYATVTICRGFG
YEVEFLPFDRRTLEIDYAVLAARLSRRPADVIYLDASSILRFIDARALRLAAPDALICLD
ASHILGLLPVAPQTLVLDGGFDSISGSTHKTFPGPQKGLLVTDSDVVAEKVAARMPFTAS
SSHSASVGLAISLEELLPHRTAYAHQVIANARALAGLLAERGFVAGGAFGHTDTHQVW
VHFPEGNTPHEWGRLLTRANIRSTSVVLPSSAAPGLRLGTQELTRWGMTETDMPVADLL
ERLLLGRGDDAETVAKEVVELARAFPGVAFV

>131

MVAPLAEVDPDIAELLGKELGRQRTLEMIASENFVPRSVLQAQGSVLTNKYAEGLPGRR
YYDGCHEVDVVENIARDRAKALFGADFANVQPHSGAQANA AVLHALMSPGERLLGLDLAN
GGHLTHGMRLNFSGKLYETGFYGVDPATHLIDMDAVRAKALEFRPKVLIAGWSAYPRILD
FAAFRSIADEVGAKLWVDMAHFAGLVAVGLHPSVPVPHADVSTTVHKTLLGGGRSGLIILGK
QEFATAINSVAVFPQQGGPLMHVIAAGKAVALKIATPEFTDRQRTLAGARILADRLTAA
DVTKAGVSVVSGGTDVHLVLDLRLNSPFDGQAAEDLLHEVGITVNRNVVNDPRPPMVT
GLRIGTPALATRGFGAEFTEVADI IATVLTGGSDVVAALRQOVTRRLARDFPLYGGLED
WSLAGR

>132

MSAPLAEVDPDIAELLAKELGRQRTLEMIASENFVPRAVLQAQGSVLTNKYAEGLPGRR
YYGGCEHVDVVENLARDRAKALFGAEFANVQPHSGAQANA AVLHALMSPGERLLGLDLAN
GGHLTHGMRLNFSGKLYENGFYGVDPATHLIDMDAVRATALEFRPKVI IAGWSAYPRVLD
FAAFRSIADEVGAKLLVDMAHFAGLVAAAGLHPSVPVPHADVSTTVHKTLLGGGRSGLIIVGK
QQYAKAINS AVFPQQGGPLMHVIAAGKAVALKIATPEFADRQRRTLSGARI IADRLMAP
DVAKAGVSVVSGGTDVHLVLDLRLNSPLDGQAAEDLLHEVGITVNRNAVNDPRPPMVT
GLRIGTPALATRGFGDTEFTEVADI IATALATGSSVDVSALKDRATRLARAFPLYDGLLE
WSLVGR

>133

TLNDSLTAFPDIAALIDGELRRQESGLEMIASENYAPLAVMQAAGGSVLTNKYAEGYPGR
RYYGGCEHVDVQVEQLAIDRVKALFGAEYANVQPHSGATANAATMHALLNPGDTILGLSLA
HGGHLTHGMRLNFSGKLYHATAYEVSKEDYLVDMDAVAEAAARTHRPKMI IAGWSAYPRQL
DFARFRAIADEVDVAVLMDMAHFAGLVAAAGVHPSVPVPHAVVTSTTHKTLGGPRGGIILC
NDPAIAKKINS AVFPQQGGPLEHVI AAKATAFKMAAQPEFAQRQRCLDGARILAGRLT
QPDVAERGI AVLTGGTDVHLVLDLRLNSPLDGQAAEDRLAAVDITVNRNAVFPDPRPPMI
TSGLRIGTPALAAARGF SHNDFRAVADLIAAALTATNDDQLGPLRAQVQRLAARYPLYPEL
HRT

>134

SMSAPLAEVDPDIAELLAKELGRQRTLEMIASENFVPRAVLQAQGSVLTNKYAEGLPGR
RYYGGCEHVDVVENLARDRAKALFGAEFANVQPHSGAQANA AVLHALMSPGERLLGLDLA
NGGHLTHGMRLNFSGKLYENGFYGVDPATHLIDMDAVRATALEFRPKVI IAGWSAYPRVLD
FAAFRSIADEVGAKLLVDMAHFAGLVAAAGLHPSVPVPHADVSTTVHXTLLGGGRSGLIIVG

KQQYAKAINS AVFPGQQGGPLMHVIAGKAVALKIAATPEFADRQRRTLSGARI IADRLMA
PDVAKAGVSVVSGGTDVHLVLDLDRSDPLDQAAEDLLHEVGITVNRNAVNDPRPPMVT
SGLRIGTPALATRGFDTEFTEVADI IATALATGSSVDVSALKDRATRLARAFPLYDGLE
EWSLVGR

>135

VRYQPLNELDPEVAAA IAGELARQRDTLEMIASENFVPRSVLQAQGSVLTNKYAEGYPGR
RYYGGCEQVDI IEDLARDRAKALFGAEFANVQPHSGAQANA AVLMTLAE PGDKIMGLSLA
HGGHLTHGMKLNFSGKLYEVVAYGVDPEMTRVMDQVREIALKEQPKVI IAGWSAYPRHL
DFEAFQSI AA EVGAKLWVDMAHFAGLVAAGLHPSVPYSDVVSSTVHKTLGGPRSGI I LA
KQEYAKKLNSSVFPQQGGPLMHVA AAKATSLK IAGTEQFRDRQARTLEGARILAERLTA
SDAKAAGVDVLTGGTDVHLVLDLDRNSQMDGQAAEDLLHEVGITVNRNAVPFDRPPMVT
SGLRIGTPALATRGFDI PA FTEVADI IGTALANGKSADIESLRGRVAKLAADYPLYEGLE
DWTIV

>136

SSVGVLLRQDPELAEI LFAEGRRQSTTLQLIAAENFTSPAVLAALGSPLANKYAEGYPGA
RHHGGCEIVDVAERLAAQRAQALFGAEHANVQSHSGSSAVLAAYAALLRP GDTV LALGLP
YGGHLTHGSPANFSGRWFDFVGYGVDAETGLIDHDQVRTLARARRPKAIVCGSIAYPRHL
DYAAFRDI ADEVGAYLIADA AHP IGLVAGGAAPSPVPYADIVCATTHKVLGRGPRGMILC
GSELAERVDRAVFPFTQGG AQMHTIAAKAVAFGEAATPAFAAYAHQV VANARALAAHLAA
EGLVVTGGTDTHLLTADPAPLGV DGTARGLLAAAGI VLDCCALPHADARGRLRGTA AV
TTQGMGEREMRAVATLVAGVLRGTTDPAAARADVRDLTAEFPPYP

>137

MKHLPAQDEQVFNAIKNERERQQT KIELIASENFVSEAVMEAQGSVLTNKYAEGYPGKRY
YGGCEHVDVVEDIARDRAKEIFGA EHVNVQPHSGAQANMAVYFTILEQGD TVLGMNLSHG
GHLTHGSPVNFSGVQYNFVEYGV DKETQYIDYDDVREKALAHKPKLIVAGASAYPRTIDF
KKFREI ADEVGAYFMVDMAH IAGLVAAGLHPNPVYADFVTTTTTHKTLRGPRGMILCRE
EFGKKIDKSI FPGIQGGPLMHVIAAKAVSFGEVLQDDFKTYAQNVISNAKRLAEALTKEG
IQLVSGGTDNHLILVLDLRS LGLTGKVAEHLVDEIGITSNKNAIPYDPEKPFVTS GIRLGT
AAVTSRGF DGDAL EEVGAI IALALKNHEDEGKLEEARQ RVAALTDKFP LYKE

>138

MKYL PQDPOVF AAIEQERKRQHAKIELIASENFVSRAVMEAQGSVLTNKYAEGYPGRRY
YGGCEYVDIVEELARERAKQLFGAEHANVQPHSGAQANMAVYFTVLEHGDTV LGMNLSHG
GHLTHGSPVNFSGVQYNFVAYGVDPE THVIDYDDVREKARLHRPKLIVAAA SAYPRIIDF
AKFREI ADEVGAYLMVDMAH IAGLVAAGLHPNPVYAHFVTTTTTHKTLRGPRGMILCQE
QFAKQIDKAI FPGIQGGPLMHVIAAKAVAFGEALQDDFKAYAKRVVDNAKRLASALQNEG
FTLVSGGTDNHL LLDLRLPQQLTGKTA EKVLDEVGITVNKNTIPYDPESPFVTS GIRIGT
AAVTRGF GLEEMDEIAAI IGLVLKNV GSEQALEEARQ RVAALTDPTS RSAAGT

>139

FDKEDYKAFDPELWNAI DAEAERQNNIELIASENVVSKAVMAAQGTLT NKYAEGYPGK
RYYGGTAVIDVETLAI ERAKKLFGAKFANVQPHSGSQANA AVYMSLIQPGD TVMGMDLS
AGGHLTHGAPVSFSGKTYNFVSYNVDKESEL LDYDAI LAQAKEVRPKLIVAGASAYSRI I
DFAKFREI ADAVGAYLMVDMAH IAGLVAAGLHPNPVYAHVTTTTTHKTLRGPRGGLILT
DDEDI AKKLN SAVFPGLQGGPLEHVIAAKAVALKEALDPAFKEYGENVIKNAAMADVFN
QHPDFRVISGGTNNHLFLVDVTKVVEN GKVAQNVLEEVNITLNKNSIPYEQLSPFKTSGI
RVGSPAITSRGMGEAESRQIAEWMVEALENHDKPEVLERIRGDVKVLTDAFPLY

>140

FDKEDFESFDPELWAAIHAE EIRQQNIELIASENIVSKAVMAAQGSVLTNKYAEGYPGK
RYYGGTEAVDVVENLAI ERAKELFGAKFANVQPHSGSQANA AAYMALIQPGD TVLGMMDLN
AGGHLTHGASVNFSGKTYHFVYPYGVNSE TELLDYDEILKIAKQVQPKLIVAGASAYSRLI
DFAKFREI ADSV GAKLMVDMAH IAGLVATGAHPNPLPYADVTTTTTHKTLRGPRGMILT
NDEALAKKINS AIFPGTQGGPLEHVIAAKAVAFKEALDPEFTTYIEQVIKNTQAMAE EFA
KVEGLRLIAGGSDNHLNLKVL DLGINGKEAQDLLDSVHITLNKEAIPDETLSPFKTS GV
RIGAAAITSRGFKEVEAKKVAQLVSEALVNHDNQEKLAEVRKAALELTRQFPL

>141

VWNRSLAETDPEIARAI ALEITRQGA KLEL IASENFVSRVLEAQGSVLTNKYAEGYPGA
RYYGGCEYVDIVESVAIRRAKEIFGAGHANVQPHSGAQANMAAYFAFLEPGDTIMGMLA
HGGHLTHGAKINFSGRYFRYVVPYGV EETGRIDYDRMHAIAREHRPKLIVGGASAYPREL
DFARMRAIADDV GALLMIDMAHIAGLIAAGLHMSVPVYADVVT TTTTHKTLRGPRGGMILC
PEEYAAAI DKAVFPGIQGGPLMHVIAAKAVALGEAQRPEFKTYQEQIVKNARALAQALQE
RGFELVAGGTDTHLILVDLRNKGLTGVAEDLLDRVDVTVNKNMVPFDPQPPRVTS GIRI
GTPAVTTRGMKEDSMVQIAEVI SLTLDHPEEGAVQARAKAIVAELCAAHPFLKL

>142

MDIEIIRKTDPEIADAIEKELIRQRNKIEL IASENFVSRVMEAMGSPLTNKYAEGYPNK
RYYGGCEYVDIAEELARERLKKLFGAEHANVQPHSGAQANMAAYFALIKPGD TVLGMDLA
HGGHLTHGSKVNFSGQIYNFVSYGVREDTGYIDYDEVERVAKKHKPKLIVAGASAYPRII
DFKRREIADKVGAYLMDMAHIAGLVAAGLHPNPVYADVVT TTTTHKTLRGPRGGAILC
KEEYAKAIDKALFPGTQGGPLMHI IAAKAVCFKEALTDEFKEYQKRIVENAKALANALME
RGINLVSGGTDNHLMLLDLRNTGITGKELETRLDEVNITCNKNAIPFDPLGPNVTS GVR
GTPAVTTRGMKPEDMVEIADIIVNVI R DENYKEKAKERVANLLKKYPLYED

>143

MRHLFNTDAEIEYEAIVKEYERQFYHLELIASENF TSLAVMEAQGSVMTNKYAEGLP HKRY
YGGCEFVDIAEDLAIERAKALFDAEHANVQPHSGTQANMAVYMAVLKPGD TIMGMDLSHG
GHLTHGAKVNFSGKIYNAVYYGVHPETHLIDYDQLYRLAKEHKPKLIVGGASAYPRVIDW
AKLREIADSVGAYLMDMAHYAGLIAGGVYPNPVPYAHFVTSTTHKTLRGPRSGFILCKK
EFAKIDKSVFPGIQGGPLMHVIAAKAVAFKEAMSQEFKEYARQVVANARVLAEEFIKEG
FKVVS GGTDSHIVLLDLRDTGLTGREVEEALGKANITVNKNVAVPFDPLPVKTS GIRLGT
PAMTTRGMKEDQMRIIARLISKVIKNI GDEKVIEWRQEVIEMCEQFPLYPELREEIN

>144

MVSTLKRDEALFELIALEEKRQREGLELIASENFVSKQVREAVGSVLTNKYAEGYPGARY
YGGCEVIDRVE SLAIERAKALFGAAWANVQPHSGSQANMAVYMALMEPGDTLMGMDLAAG
GHLTHGSRVNFSGKLYKVVS YGVRPDT ELIDLEEVRRLALEHRPKVIVAGASAYPRF WDF
KAFREIAD EVGAYLVVDMAHFAGLVAAGLHPNPLPYAHVVTSTTHKTLRGPRGGLILSND
PELGKRIDKLI FPGIQGGPLEHVIAAGKAVAF FEALQPEFKYSRLVVENAKRLAEELARR
GYRIVTGGTDNHLFLVDLRPKGLTGKEAEERLDAVGITVNKNAI PFDPKPPRVTS GIRIG
TPAITTRGFTPEEMPLVAELIDRALLEGPSEALREEVRR LALAHMP

>145

TNFDFLAQGDPAIAAII GRELQRQEHLELIASENFASPAVMAAQGSVLTNKYAEGLP SK
RYYGGCEFVDQAEELAIERAKELFGAAHANVQPHSGAQANFAVFLTL LQPGDTFLGMDLS
HGGHLTHGSPVNVSGKWFNAGHYGVNRETERLDYDAIRELALQHRPKL IICGYSAYPRTI
DFAKREIAD EVGAYLLADMAHIAGLVAAGLHPSPIPHCDVVT TTTTHKTLRGPRGGLILT
RDAELGKKLDKSVFPGTQGGPLEHVIAAKAVAFGEALRPEFKTYS AQVIANAQALARQLQ
ARGLKIVSDGTDNHL LLDLRSIGMTGKVADLLVSDVNI TANKNTVPFDPE SPFVTSGIR
LGTAAMTRGFKEAEFAIVADIIADRLLNPEDSSMEDSCR RRVLELCQRFFLYPHLS

>146

THIDWL VQTDPLVAEMVQREVQRQQHLELIASENF TSPAVMAAQGT VLTNKYAEGLP GK
RYYGGCEFVDEVEQLAIDRAKELFGAAHANVQPHSGAQANFAVFLALLNPGDTIMGMDLS
HGGHLTHGSPVNVSGKWFNVVHYGVHPETERLDMDQVRDLARQHRPKL IICGYSAYPRVI
PFAEFRQIAD EVGAYLMADIAHIAGLVASGYHPNPVPLCDVVT TTTTHKTLRGPRGGLILT
RDEDLGKKLDKAVFPGTQGGPLEHVIAAKAVAFGEALKPEFKAYCGQVIRNAQALAAGLQ
ARQLRLVSGGTDNHLMLIDLRSVNL TGKEADRLMGEIHI TTNKNTIPFDPASPFVTS GLR
LGT PALTTRGFTEVEFAEVAEII SDR LHAPEDA IKNRCRERVAALCAQFPLYPHLQ

>147

FNNDP LQKYDKELFDLLEKEKNRQIETINLIASENL TNTAVRECLGDRI SNKYSEGYPHK
RYYGGNDYVDKIEELCYKRALEAFNVSEEEWGVNVQPLSGSAANVQALYALVGVK GKIMG
MHLCSGGHLTHGFFDEK KVSITSDLFESKLYKCNSEGYVDMESVRNLALS FQPKVICG

YTSYPRDIDYKGFREICDEVNAYLFADISHISSFVACNLLNNPFTYADVVTTHKILRG
PRSAIIFFNKRNPGIDQKINSSVFPFQGGPHNNKIAAVACQLKEVNTPEFKEYTKQVL
LNSKALAECLLRNLDLVTNGTDNHLIVVDLRKYNITGSKLQETCNAINIALNKNTIPSD
VDCVSPSGIRIGTPALTRGCCKEKMDEFIADMLLKAILLTDLQKYGKKLVDFKKGLVN
NPKIDELKKEVVQWAKNLPFA

>148

MFNNEPLEQIDKELHDI LADEEKQRRETINLIASENLNGAVRECLGNRVSNKYSEGYPK
KRYGGNDFIDKIEELCQKRALEAFNVSDDEWGVNVQPLSGSAANVQALYALVGKGM
GMHLCSGGHLTHGFFDEKKVSITSDMFESKLYKCNQGYVDLDAVREMAISFKPKVIIC
GYTSYPRDIDYQQRQICDEVNAYLFADISHISSFVACNILLNNPFLHADVVTTTHKILR
GPRSAIIFFNKRNPGIEQKINSAVFPFQGGPHNNKIAAVACQLKEVHSPAFKEYTQQV
LLNSKALAKALISKQIDLVTNGTDNHLIVVDLRKFSITGSKLQETCNAINVSLNKNTIPSD
DVDCVSPSGVIRIGTPAMTRGAKKDEMEFIADVLARAIKITVDLQEQYGKKLVDFKKGLP
GNAQLQQLKQEVVTVWAGALPFP

>149

MPYTLSDAHHKLITSHLVDTDPEVDSIIKDEIERQKHSIDLIASENFTSTSVFDALGTPL
SNKYSEGYPGARYYGGNEHIDRMEILCQQRALKAFHVTPDKWGVNVQTLGSPANLQVYQ
AIMKPHERLMGLYLPDGGHLSHGYATENRKISAVSTYFESFPYRVNPETGIIIDYDTLEKN
AILYRPKVLVAGTSAYCRLIDYKRMREIADKCGAYLMVMDMAHISGLIAAGVIPSPEYAD
IVTTTTHKSLRGPRGAMIFFRRGVRSTDPKTGKEVLYDLENPINFVFPGHQGGPHNHTI
AALATALKQAAATPEFKEYQTQVLKNAKALESEFKNLGYRLVSNGTDSHMLVLSLREKGV
GARVEYICEKINIALNKNSIPGDKSALVPGGVIRIGAPAMTRGMGEEDFHRIVQYINKAV
EFAQQVQQLPKDACRLKDFKAKVDEGSDVLTWKKIYDWAGEYPLAV

>150

MATYALSESHKELMEVHLADFPEIAEIIKKEIQRQRESILLIASENVTSTRAVFDALGTP
MSNKYSEGYPGARYYGGNQHIDAVELTCQARALKVFNLDPEKGVNVQTLGSPANLQVY
QALMKPHDRMLGLDLPDGGHLSHGYQTPQRKISAVSTYFETFPYRVNSETGIIIDYDTLEA
NAQLYRPKILVAGTSAYCRLIDYARMRKIADSVGAYLVMDMAHISGLIAAGVIPSPEYAD
DVVTTHKSLRGPRGAMIFFRRGVRSTDPKTGKEIYDLEGPINFVFPGHQGGPHNHTI
ITALAVALKQASTPEFRQYQEQTIKNAKALEVAFKEYGYKLVDGTDSHMLVLDLRPNGI
DGARVETVLEQINIACNKNVPGDKSALSPGGIRVGAPAMTRGLGEEDFKRVVGYIDQA
IKISKSIQASLPKEANKLKDFKAKASSETIPEILNLRKEISAWASTFPLPV

>151

MATYALS PAHRNQMEVSLVDSPEIAQIMEKEIQRQRESIVLIASENFTSHAVFDALGSP
MSNKYSEGYPGARYYGGNQHIDAIELTCQARALKAFNLDPAKWGVNVQCLSGSPANLQVY
QALMRPHDRMLGLDLPDGGHLSHGYQTPARKISAVSTYFETFPYRVNLETGIIIDYDQLEA
NAELYRPKCLVAGTSAYCRLIDYARMRKIADKVGAYLIVDMAHISGLIAAGVIPSPEYAD
DVVTTHKSLRGPRGAMIFFRRGVRSTDPKTGKDIMYDLEGPINFVFPGHQGGPHNHTI
ITALAVALKQVDTPEFRQYQQQVIKNAKALEEEFKALGHKLVS DGTDSHMLVLDLRNKSL
DGARVEAVLEQINIACNKNVPGDKSALTPCGIRIGAPAMTRGMGEEDFKRIARYIDQS
INLCKVQSELPEANKLKDFKAKVASETVPEILSLRKEVAEWASTFPLPV

>152

HSNAAQTQTGEANRGWTGQESLSDSDPEMWELLQREKDRQCRGLELIASENFCSTRAALEA
LGSCLNKYSYEGYPGARYYGGAEVVDIEELLCQRRALEAFDLDPAQWGVNVQPYSGSPAN
LAVYTALLQPHDRIMGLDLPDGGHLSHGYMSDVKRISATSIFFESMPYKLNPKTGLIDYN
QLALTARLFRPRLI IAGTSAYARLIDYARMREVCDEKHAHLLADMAHISGLVAAKVIPS
FKHADIVTTTTHKTLRGARSLIFYRKGVKAVDPKTGREIPYTFEDRINFVFPVSLQGGP
HNHIAAVALKQACTPMFREYSLQVLKNARAMADALLERGYSLVSGGTDNHLVLDLR
PKGLDGARAERVLELVSI TANKNTCPGDRSAITPGGLRLGAPALTSRQFREDDFRVVDF
IDEGVNI GLEVSKTAKLQDFKSFLLKDSSETSQR LANLRQ RVEQFARAFMPGDFDEH

>153

PRDAALWSSHEQMLAQPLKSDAEVYDIKKESENQRVGLGLIASENFASTRAVLEALGSC
LNNKYSEGYPGARYYGGTEHIDELETLCQKRALQAYGLDPQCWGVNVQPYSGSPANFAVY

TALVEPHGRIMGLDLPDGGHLTHGFMTDKKKISATSIFFESMAYKVNPDYGYIDYDRLEE
NARLFHPKLI IAGTSCYSRNLDYGR LRKIADENGAYLMADMAHISGLVVAGVVPSPFEHC
HVVTTHKTLRGC RAGMI FYRRGVR SVDPKTGKEILYNLESLINSAVFPGLQGGPHNHA
IAGVAVALKQAMTPEFKEYQRQVVANCRALSAALVELGYKIVTGGSDNHLILVDLRSKGT
DGGRAEKVLEACSIACNKNTCPGDKSALRPSGLRLGTPALTSRGLLEKDFQKVAHF IHRG
IELTVQIQDDTGPRATLKEFKEKLAGDEKHQRAVRALRQEVESFAALFPLPGLPGF

>154

DRDATLWASHEKMLSQPLKDSDAEVYSI IKKESNRQRVGLEL IASENFASRAVLEALGSS
LNNKYSEGYPGQRYGGTEFIDELEMLCQKRALQAYHLDPQCWGVNVQPYSGSPANFAVY
TALVEPHGRIMGLDLPDGGHLTHGFMTDKKKISATSIFFESMPYKVYPETGYINYDQLEE
NASLFHPKLI IAGTSCYSRNLDYARLRKIADNGAYLMADMAHISGLVAAGVVPSPFEHC
HVVTTHKTLRGC RAGMI FYRKGVR SVDPKTGKETYYELES LINSAVFPGLQGGPHNHA
IAGVAVALKQAMTTEFKIYQLQVLANCRALSDALTELGYKIVTGGSDNHLILMDLRSKGT
DGGRAEKVLEACSIACNKNTCPGDKSALRPSGLRLGTPALTSRGLLEEDFQKVAHF IHRG
IELTLQIQSHMATKATLKEFKEKLAGDEKIQSAVATLREEVENFASNFSLPGLPDF

>155

DADLWSSHDAMLAQPLKDSDEVYNI IKKESNRQRVGLEL IASENFASRAVLEALGSCIN
NKYSEGYPGQRYGGTEFIDELETLCQKRALQAYKLDPCWGVNVQPYSGSPANFAVYTA
LVEPHGRIMGLDLPDGGHLTHGFMTDKKKISATSIFFESMPYKVNPDYGYINYDQLEENA
RLFHPKLI IAGTSCYSRNL EYARLRKIADENGAYLMADMAHISGLVAAGVVPSPFEHCHV
VTTTHKTLRGC RAGMI FYRKGVR SVDPATGKEILYNLESLINSAVFPGLQGGPHNHAIA
GVAVALKQAMTLEFKVYQHQQVVANCRALSEALTELGYKIVTGGSDNHLILVDLRSKGTDG
GRAEKVLEACSIACNKNTCPGDRSALRPSGLRLGTPALTSRGLLEKDFQKVAHF IHRGIE
LTLQIQSDTGVAATLKEFKERLAGDKYQAAVQALREEVESFASLFPLPGL

>156

FFNSSVHDTDPLIAQALDDERARQKNQIEL IASENIVSQAVLDALGHEMTNKTLEGYPGN
RFHGGGQFVDVVEQAAIDRAKQLFNCGYANVQPHSGTQANLAVFFLLVKPGDRILSLDLA
AGGHL SHGMKGNLSGRWF EAHNYNVDPQNEVINYDEMERIAEEVKPKLLITGGSAYPREL
DFARMAQI AKKVGAFMVDMAHIAGLVAGGAHPSPPHADIVTCTTTKTLRGPRGGLIILT
NNEEWYKQLQAVFPGVQGSLSHNVLAAKAICLGEALRPEFRDYVAQVVKNAKVLAEETLT
SRGIRIVSGGTDTHIVLDDLSSKGLNGKQAEDALARANITSNKNPIPNDSRPAEWVGM
LGVSAAATTRGMKEDEFKRLGNVVADLLEAESAGNGPEAAEKAKVTVRELTEAFPVYAH

>157

YFNTPVHERDPLVAQALDNERKRQDQIEL IASENIVSRAVLDALGHEMTNKTLEGYPGN
RFHGGGQFVDVVEQAAIDRAKELFGCAYANVQPHSGTQANLAVFFLLKPGDKVLSLDLA
AGGHL SHGMKGNLSGRWFESHNYNVDPETEVIDYDEMERIAEEVRPTLLITGGSAYPREL
DFERMGKIAKKVGAWFLVDMMAHIAGLVAGGAHPSPPHADIVTCTTTKTLRGPRGGLIILT
NNEAWFKQLQSAVFPGVQGSLSHNVLAAKAVCLGEALRPDFKVYAAQVKANARVLAETLI
ARGVIRIVSGGTDTHIVLVDLSSKGLNGKQAEDLLARANITANKNPIPNDSRPAEWVGM
LGVSAAATTRGMKEDEFRTLGTVIADLIEAEAAGNADGVVEGAKAKVATLTAAFPVYAH

>158

IGWDALTRKDPALGALLDAEAAAYQTAHLSMVASASVAAPSVLCAHGTVMSNVTAEGYPGA
RYHPGA VHFDEIERL AVERARHVFGATYANVQPHSCSSANLAVLGALLPRGGRILALDLG
SGGHLTHGADASVSGKYVDVAHYGVEPDGRLDYDRILDRA LTHRDPVVIAGASAYPRWID
FSRFRAVADRVGAWLLADISHIAGLVAAGLHPSVPDHAHATTTSTYKQLGGPRGGLIILG
KDHRA TAPDGRTELWKL MQRAVFPGFQGTDPDSAIAAKARALAIASGGAFQATMKRVVDD
AKALADALVELGYDVL TGGTDNHMVLDVRAAGLTGIVAERALEECGILVNRNKIPGDPL
PPGVGGGLRFGTNILAQRGMGPFVRECARIVHNVLS SIQAGDATHYELDGALRDAVRDR
VRALCARFPLQFDDE

>159

MKGPSEFFPREDFASDNHAEAHPAVLAALQAAAHPAPAYGGDRWTA AAEAGLRDLAGPAA
VPLL VFNATAANV LGLSVLLGPGDAI ICAEHS HLAMDECGAVEAMLGRKLLPVPAPDGKL
TPAAIEALLQRCTDVRTPRPSTVALTQATELGTCYTLAE LAELREFCDRRGLRIHMDGAR

LANAVAHLNCEVADIARFADALSFGFTKNGALGVDALLMTPGTAERAPYLRRQQGQLAS
KMRFLAAQVTALVDGGLWLDNARHANGMAGHLAAELEPLPGLRLTQRVQSNVVFVLPRA
AADSLRERWAFHLSPTGEARLMTSHATGRTDVSDLVAEVKTCLKQ

>160

MLVASSVAPPSVLACAGSSLANLTMEGYPGRRYHAGAEVADEVERLALDRARALFGARA
ANVQPHSGSSANLAVLTGLLEPGDTILGLGLDCGGHLSHGSAASVTGRYFRSVHYRTGAD
DLLDYEQIAGLAGEHRPKVIIAGASAYPRQIDFARFREIADTVGAYLIADISHIAGLVAA
GLHPSVPDHAHLTTTSTYKQLYGPRGGLILLGRDADAPGPDGRLPLRKLADRAVFLVQG
TPDLGNVAAKARALHAAAQPAFRALMARVLDTATAIAEQLDRRGLCVITGGTDTHMVLAD
LRPEHVSAGADAEAEACGILVNRNRVPGDDTTPRVTTGGIRIGTNTLAARGVSPETAQQC
AELVADVVEELRTAGAVHPSTIAKIRAEVRRICAEHPVPGYAKPRI

>161

MTVGAGGKTSADADPLMLVRAIADARRAAHALNLVPSENRISPLASLPLASDFYNRYFF
NTDGDPLFWEFRGGEDIAHIEALGAAALRRMASARYCNVRPISGMSAMILTVAALSPPGS
TVVSDQNSGGHYATPALLGRLGRRSRLLNCKDGEVDESELAEVLAPGDVALVYVDVQNC
VRVPDFRRMSDVIREVSPGTRLYVDASHYLGVLGGLLANPLDCGADAFGGSTHKSFPGP
HKGVI FTNAEDVDESLSAQFDLVSSHFAETLALS LAALEVEDRMGDYARATNDNARRL
AGALADAGFRVYGD SATGYTDTHQVWVELDGVAAAYALSNRLAEGGIRVNLQSSMPGMSG
VHLRLGSNEVTFEGAGPQAIEELAGALVTARERALGPRTVHEIRGRFGAPFYTDPEKLV
EAGL

>162

LLALLGEIEKEQRINEAAVNLPSENRI SPWAGAPLRTDFYNRYFFNDSLDPQGWQFRGG
EGIGRLEKELALPALRRLGRADHVNI RPVSGMSAMLVLLGLGGE PGDGVVCVDAETGGH
YATGRQIAMLGRRPLPVRVAVGRVDLALRTALTSCHVPLVYLDLQNSLWELDVAGVAEV
IARTSPRTVLHVDCSHTLGLILGGSHKNPLDLGADTTGGSTHKTFFPGPKGVLFTRDENL
SRKIRDAQFFTISSHHFAETLALALAAAEFEHFGAAYSQVVLINARAFHRLRERGFVGV
EGGPQLTDTHQVWVRLPLEESADAFSAQLASLGIRVNVQTELPDIPEPALRLGVSEITLN
GGREPAMETLAEIFALVRAGEATKAVDLFQVLPHEMGEPYFFTGLP

>163

MTVGADGKTAADVDPMLLGRAIVDADRRAAHALNLVPSENRISPLAALPLGSDFYNYRYFF
NTAGDPLFWEFRGGEDIAHIEALGAQALRRMASAQYCNVRPISGMSAMILTVAALSAPGR
TVVSDQNSGGHYATPALLGRMGRHSRLLGCKDQVDESELADVLAPGDVLDVYVDVQNC
VRVPDFGLMADVNVQSPATRLYVDASHYLGVLGGLVENPLACGADAFGGSTHKSFPGP
HKGVI FTNAEDVDESLSAQFDMVSSHFAETLALSLSALEVEPRIGDYAWATTDNARRL
ACALADAGFRVYGD SRSGYTDTHQVWVELDGTADAYALSNRLAEAGIRVNLQSSMPGMSG
VHLRLGSNEVTFEGAGPQAIEQLAGALATARERALEPRTVSEIRGRFGAPFYTDPEKLV
EAGL

>164

MTVRAAGKTSADADPLALARA IADADRRASALNLVPSENRVSPLASLPLASDFYNRYFF
NTDGDPLFWEFRGGEDIAHIEALGAAALRRMAMARYCNLRPISGMSAMILTVAALSKPGS
TVVSDQNSGGHYATPALLGRLGRRSRLLTCKDGA VDESELADVLAPGGVDLVYVDVQNC
VRVPDFRLMSDVIRNVSPGTRLYVDASHYLGVLGGLVDNPLDCGADAYGGSTHKSFPGP
HKGVI FTNAEDVDESLSAQFDLVSSHFAETLALS LAALEVEDRIGDYARATNDNARRL
AGALAEAGFRVCGDTGTGYTDTHQVWVELAGTDEAYALSNRLAEAGIRVNLQSSMPGMSG
VHLRLGSNEVTFEGAGPQAIEELAGALVQARERALGPHTVSEIRGRFGAPFYTDPEKLV
EAGL

>A

MNARKAPEFFPAWPQYDDAERNGLVRALEQQGWWRMGDEVNSFEREFAAHGAAHALAVT
NGTHALELALQVMGVGPGTEVIVPAFTFISSSQAAQRLGAVTVPVVDVDAATYNLDPEAVA
AAVTPRTKVIMPVHMAGLMADMDALAKISADTGVPLLQDAHAHAGARWQKRVGELDSIA
TFSFQNGKLMTAGEGGAVVFPDGETEKYETAFLRHSCGRPRDDRRYFHKIAGSNMRLNEF
SASVLRACLARLDEQIAVRDERWTL S RLLGAIDGVVPQGGDVRADRNSHYMAMFRI PGL
TEERNALVDRLVEAGLPAAFAAFRAIYRTDAFWELGAPDESVDIAIRRCPNDAISSDCV

WLHHRVLLAGEPELHATAEIIADAVARA

>B

MGSSPDAGIDFPAPWQHDDAERAALLRALDQGGQWWRVGGSEVDEFEREF AEYHGAGHALA
VTNGTHALELALQVLDVGPTEVIVPAFTFISSSQAVQRLGAVAVPVDVDPD TYCLDVAA
AEDAVTSRTSAIMPVHMAGQFADMDRLDKLSASTGVPVQDAAHAHGAHWRGKRVGELGS
IATFSFQNGKLMTAGEGGAVLFADQAQWEKAFVLHSCGRPKGDRGYFH LTSGSNFRMNEF
SAAVLRALQRLDSQIATRQARWPVLSALLAGIDGVVPQTVDPDRSDRNPSYMMAMFRMPGV
TEERRNAVVDLVRGIPAFMAFRAVYRTQAFWETGAPDLTPEELAARCPVSEEITRDCV
WLHHRVLLGAEQVRRLAADVADVVAGA

>C

MNARQAPEFPRWPQYDETERDGLIRALEQGGQWWRMGGDEVDSFEREF AEHHGAPHALAVT
NGTHALELALQCLGVGPTEVIVPAFTFISSSQAAQRLGAVAVPVDVHPD TYCIDAAAVA
AAVTPRTRAIMPVHMAGLIADMDALAKISADTGVPLLQDAAHAHGARWQ GKRVGELGSVA
AFSFFQNGKLMTAGEGGALLFPDELYEAAFLRHSCGRPRDRRYLHRTAGSNLRMNEFSA
AVLRAQLSRLDQQIALREERWTLLSRLAAIDGVVPQGGDVRVDRKSHYMMAMFRV PGLGE
ERRNALVDRLVEAGLPFAAFRAIYRTDAFWEIGAPDETPDAIAARCPHTEAISQDCVWL
HHRVLLAGEAEMHATAEIIADLVARA

>D

MLDKELDLDFPAPWQYDDTERAGLVRALQGGQWWRIGGSEVDEFEREF AEANGAPHALAV
TTGTHALELALQVLDVGPTEVVVPAFTFISSSQAAQRLGAVAVPVDVLD TYCLDPEAV
AAAITPRTAAIMPVHMAGQICDMDALGKLSADSGVPLLHDAHAHGGRRWDQGV SALGTM
AAFSFQNGKLMTAGEGGAVTFPDSEQYETAFLRHSCGRPRTDRTYRHQ TSGSNFRMNEFT
ASVLRALQARLDGQIDTREQRWPVLAGQLARITGVLPQATDDRCTRNP HYMMAMFRVPGIS
EQQRNELVDALVARGLPFAAFRAIHRTKAFWETGAPDEPVEAIARRCPNSEALSTDCVW
LHHRVLLGTEEQMHAVEVSDALAAS

>E

MTPTSGDDVLSFSPWPQHGAEEERAGLLRALDQKGWWRDAGQEVDLFEREF ADHHGAPHAI
ATTNGTHALELALGVMGIGPGDEVIVPAFTFISSSLAVQRMGAVVPADVRP DTYCLDAD
AAAALVTPRTKAIMPVHMAGQFADMDALEKLSVATGVPVLQDAAHAHGAQW GRRVGE LG
SIAAFSFFQNGKLMTAGEGGALLLPDESFEAFLQHCCGRPPGDRVYRH LTOGSNYRMNE
FSASVLRALQRLKLDQLRIREERWAQLRTALAAIDGVVPQGRDERGDLHSHYMMAMVRLPG
ISARRRALVDALVERGVPAFVGFPPVYRTEGFARGPAPADAEELAKSCPVAEEIGSDCL
WLHHRVLLADVTTLDRLAEVFSGLVGAL

>F

MSSGVQLGSAFRVWPQYDDAERTGLIRALEQGGQWWRMGGGEVERFEREF AEYHGGEHALA
VTNGTHALELALQVLDVGPTEVIVPAFTFISSSQAAQRLGAVAVPVDVDP ETYCIDPAE
AAKAITPRTRAIMPVHMAGQLADMDALEKVAADSGVPLIQDAAHAQGATWNGRR LGELGS
VAAFSSFFQNGKLMTAGEGGAVLFPTEMAEHAFLRHSCGRPRNDRGYFH RTSGSNFRMLNEF
SASVLRALQARLDGQIRITREERWPLLSLLAEIPGVVPQRLDRRPRDRN PHYMMAMFRVPRI
TEERRARVVDTLVERGVPAFVAFRSVYRTDAFWEMGAPDLSVDELARLPPLRGLTTDCLW
LHHRVLLGTEEQMHEVA AVIADVLGS

>G

MNARPAPEFPTWPQYDDEERTGLIRALEQGGQWWRMGGEEVSSFEQEFAD FHGAPHAFAVT
NGTHAFELALQVMGAGPGTEVIVPAFTFISSSQAAQRIGAVAVPVDVDPD TYNIDVAAAA
AAVTPRTRVIMPVHMAGLMADMDALGKLSADTGVAIILQDAAHAQGARWQ GKRVGELGTVA
AFSFFQNGKLMTAGEGGAVLFPENDLYEAAFLRHSCGRPRTDRHYKHQVAGTNMRLNEFSA
AVLRAQLRRLPAQTELRDRRWALLSRLLGAVDGVVPQGGDVRADQNTHYMMAMFRIPGITE
ADRNTLVDRLEAGLPFAAFRASTAPTSPGENRRPRRRPWSSVAERCPHSRGHQORTAS
GCTTGSSSPTSGTWNNSAAEIIADAVAAI

>H

MGSSPDAGIDFPAPWQHDDAERAALLRALDQGGQWWRVGGSEVDEFEREF AEYHGAGHALA
VTNGTHALELALQVLDVGPTEVIVPAFTFISSSQAVQRLGAVAVPVDVDPD TYCLDVAA

AEDAVTSRTSAIMPVHMAGQFADMDRLDKLSASTGVPVVQDAAHAHGAHWRGKRVGELGS
IATFSFQNGKLMTAGEGGAVLFADQAQWEKAFVLHSCGRPKGDRGYFHLLTSGSNFRMNEF
SAAVLRALQGRLLDSQIATRQARWPVLSALLAGIDGVVPQTVDPDRSDRNPSPYMMAMFRMPGV
TEERRNAVVDDELVRRGIPAFMAFRAVYRTQAFWETGAPDLTPEELAARCPVSEEITRDCV
WLHHRVLLGAEEQVRRLAAVVADVAGA

>I

MNARRTPEFPTWPQYDDGERTGLIRALEQQQWWRMGGSEVDSFEGEFADFHGAPHALAVT
NGTHALELALQCLGVPTEVIVPAFTFISSSQAAQLGAVAVPVDVLDLTYNIDVAAAA
SAVTPITKAIMPVHMAGLIADMDALGELSADTGVPLLQDAAHAHGARWQKRVGELGTVA
SFSFQNGKLMTAGEGGALLLPDEETYEAFLRHSCGRSRTDRRYMHQTAGTNMRLNEFSA
AVLRAQLGRLLDAQITLRDQRWTLRSRLLEIDGVVPQGSDFRDRNSHYMMAMFRIPGISE
EARNALVDLVEAGLPFAAFRAIYRTDAFWETAAPDPTVVKLAESCPHTEAISTDCIWL
HHRVLLASEEALHTTAEIADAVAAR

>J

MRLRSELPAPWPQYGDDEEREALIRALDQQQWWRIGGGEVDAFEAEFAAAHGSEHALAVTNG
THALELALVGLVGDSEVIVPAFTFISSSQAAQLGAVAVPVDVDPDTCIDPSAVEAA
IGPKTRAIMPVHMAGQMCMDALGKLSADSGVPLIQDAAHAHGARWRGQKRVGELGSVAAF
SFQNGKLMTAGEGGAVLFPDAEMYERGFVRHSCGRPRTRDRGYFHRTSGSNFRLNEFSASV
LRAQLTRLDGQITTREQRWPVLSRLLAIEIPGVVPQSRDDRDRNPHYMMAMFRVPGITEER
RAKVVDTLIERGVPAFVAFRAVYRTDAFWEVAAPDLTVDELARRCPHSEALTRDCLWLHH
RVLLGSEEQMHEVAAVVADVLAGA

Appendix 3:

Publication

ARTICLE

Received 13 Oct 2016 | Accepted 15 May 2017 | Published 26 Jun 2017

DOI: 10.1038/ncomms15935

OPEN

An L-threonine transaldolase is required for L-threo- β -hydroxy- α -amino acid assembly during obafluorin biosynthesis

Thomas A. Scott¹, Daniel Heine¹, Zhiwei Qin¹ & Barrie Wilkinson¹

β -Lactone natural products occur infrequently in nature but possess a variety of potent and valuable biological activities. They are commonly derived from β -hydroxy- α -amino acids, which are themselves valuable chiral building blocks for chemical synthesis and precursors to numerous important medicines. However, despite a number of excellent synthetic methods for their asymmetric synthesis, few effective enzymatic tools exist for their preparation. Here we report cloning of the biosynthetic gene cluster for the β -lactone antibiotic obafluorin and delineate its biosynthetic pathway. We identify a nonribosomal peptide synthetase with an unusual domain architecture and an L-threonine:4-nitrophenylacetaldehyde transaldolase responsible for (2S,3R)-2-amino-3-hydroxy-4-(4-nitrophenyl)butanoate biosynthesis. Phylogenetic analysis sheds light on the evolutionary origin of this rare enzyme family and identifies further gene clusters encoding L-threonine transaldolases. We also present preliminary data suggesting that L-threonine transaldolases might be useful for the preparation of L-threo- β -hydroxy- α -amino acids.

¹Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. Correspondence and requests for materials should be addressed to B.W. (email: barrie.wilkinson@jic.ac.uk).

Despite the great structural diversity of biologically important natural product scaffolds, a number of privileged structural motifs occur within them with high frequency. An important example is found in the β -hydroxy- α -amino acids which are constituents and intermediates in the biosynthesis of many agriculturally and medicinally valuable bioactive natural products. These include the antibiotics chloramphenicol¹ and vancomycin², sphingofungin antifungal agents³ and the proteasome inhibitor salinosporamide⁴ among others. They are also of great importance as chiral building blocks for chemical synthesis and *l*-threo-3,4-dihydroxyphenylserine (Droxidopa) is itself a (pro)drug used in the treatment of Parkinson's disease⁵. Although many different synthetic methods have been reported for their chemical synthesis—including methods based on glycine enolates⁶, glycinamides⁷, glycine Schiff's base⁸ and the aminohydroxylation of olefins⁹—the production of β -hydroxy- α -amino acids by enzymatic means is particularly attractive as these can set two stereocentres in a single reaction that can be performed in a one-pot process with minimal protection of substrates and under mild, aqueous conditions. One such example is found in the threonine aldolases (TAs) although their utility is limited due to low synthetic yields and modest diastereoselectivity (they are highly stereoselective for the α -carbon)^{10–12}. As such the discovery and characterization of new enzymes with utility for the synthesis of β -hydroxy- α -amino acids is desirable to expand our toolkit for asymmetric synthesis. Given the presence of β -hydroxy- α -amino acid derived moieties in many natural product structures we decided to target biosynthetic gene clusters (BGCs) encoding for the production of such molecules.

On this basis the antibacterial agent obafluorin^{13,14} (**1**; Fig. 1) produced by *Pseudomonas fluorescens* ATCC 39502 attracted our attention as classical feeding experiments suggested its β -lactone moiety derives from cyclization of an unusual β -hydroxy- α -amino acid intermediate (2*S*,3*R*)-2-amino-3-hydroxy-4-(4-nitrophenyl)butanoate (**2**) via homologation of *l*-4-aminophenylalanine (**3**)^{15–18}. Previously, Herbert and Knaggs proposed a biosynthetic pathway in which **2** derives from 4-aminophenylpyruvate (**4**; derived from transamination of **3**) and glyoxylate via a decarboxylative, thiamine diphosphate (ThDP)-dependent mechanism as shown in Supplementary Fig. 1 (ref. 16). If correct, such an enzyme, when coupled with appropriate aminotransferases, would provide access to valuable substituted β -hydroxy- α -amino or α -hydroxy- β -amino acids in enantiomerically enriched form.

Moreover, the biosynthesis of **1** is of additional interest as the β -lactone structural motif occurs infrequently in nature but is usually associated with molecules possessing potent and valuable biological activity^{19,20}. For example, **1** exhibits narrow spectrum antibacterial activity^{13,14}, despite being a substrate for β -lactamases and sensitive to acidic conditions, undergoing facile hydrolysis or ring opening in the presence of nucleophiles. It protects mice infected with a clinical isolate of *Streptococcus pyogenes* when dosed systemically, and microscopic examination of *Escherichia coli* grown at sub-lethal doses showed unusual cell elongation¹⁴. Taken together these observations suggest that **1** acts in a specific manner rather than as a general acylating agent (β -lactones are generally effective electrophiles able to form reversible covalent linkages with nucleophilic residues of target proteins). The molecular target of **1** remains elusive making studies on its biological activity and biosynthesis of particular interest: in an era of extensive antimicrobial resistance the identification of new antibacterial targets is of great potential significance.

Using genome mining, in conjunction with mutational and biochemical analysis, we report the pathway for the biosynthesis of **1** and show that the BGC includes genes involved in the

biosynthesis of the precursors **4** and 2,3-dihydroxybenzoic acid (**5**). These co-localize with additional genes including those encoding for a ThDP-dependent phenylpyruvate decarboxylase (pPDC) ObaH and a putative serine hydroxymethyltransferase (SHMT)/*l*-TA ObaG, both of which are essential for the biosynthesis of **2** and therefore, ultimately, **1**. Our data show that ObaG actually encodes for an *l*-threonine transaldolase (*l*-TTA), and we propose that this group of enzymes may offer utility as tools for the synthesis of *l*-threo- β -hydroxy- α -amino acids.

Results

Identification of the **1 BGC.** The genome of *P. fluorescens* ATCC 39502 was sequenced at the Earlham Institute (Norwich, UK) using the Pacific Biosciences (PacBio) RSII platform, and assembly using the HGAP2 pipeline gave a single circular contig of ~6.15 Mb. The sequence was submitted to the open access genome mining platform antiSMASH 3.0 (ref. 21) which predicted sixteen BGCs, seven of which included nonribosomal peptide synthetase (NRPS) encoding genes. Functional assignments were made for each gene by comparing the deduced amino acid sequence with proteins of known function which allowed us to identify one BGC encoding the activities we anticipated would be required for the biosynthesis of **1** (Fig. 1; Supplementary Table 1). The *oba* BGC sequence has been deposited with GenBank under the accession no. KX931446.

To verify the *oba* locus we subjected the BGC to mutational analysis (see below). To facilitate genetic modification we generated pT51 (Supplementary Fig. 2 and Supplementary Table 2; GenBank accession no. KX931445), a variant of the suicide vector pME3087 (ref. 22) which was first sequenced *de novo* by primer walking (Supplementary Table 2; GenBank accession no. KX931444). pT51 incorporates an expanded multiple cloning site and the *sacB* gene which allows for positive selection of gene replacements after secondary recombination and plasmid loss²³ (see Supplementary Note 1 for construction). Gene inactivation experiments were performed by generating in frame deletions using pT51, and all resulting mutants were verified by PCR amplification across the deleted genomic region followed by sequencing of the amplicon; at least three independent mutants were checked. All mutants were functionally confirmed and checked for the lack of polar effects by genetic complementation through ectopic expression of the deleted gene under the control of the *tac* promoter using the expression plasmid pJH10TS, based on pJH10 previously reported by Thomas and co-workers²⁴ (Supplementary Note 2 and Supplementary Table 2). As required, chemical complementation with putative pathway intermediates was carried out in parallel by the addition of exogenous material to growing cultures. All strains and their complemented derivatives were grown in triplicate under **1** producing conditions (alongside the wild type (WT) strain), extracted with ethyl acetate and the extracts subjected to high performance liquid chromatography (HPLC) and high performance liquid chromatography-mass spectrometry (HPLC-MS) analysis.

Bioinformatics and mutational analysis of the **1 BGC.** Bioinformatics analysis suggests that three genes (*obaJLN*) encode for the well understood pathway to **5** from chorismate^{25,26} and deletion of either *obaJ* (an isochorismatase) (Supplementary Fig. 3) or *obaL* (a 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase) (Fig. 2a) led to the loss of **1** production and 4 to 5-fold elevated levels of 4-nitrophenylethanol (**6**) and 4-nitrophenylacetic acid (**7**), which are proposed shunt metabolites of the biosynthesis of **2** (ref. 17). The addition of exogenous **5** was able to restore the biosynthesis of **1** (to 71% (Δ *obaJ*) and 68% (Δ *obaL*) of WT titres), as was

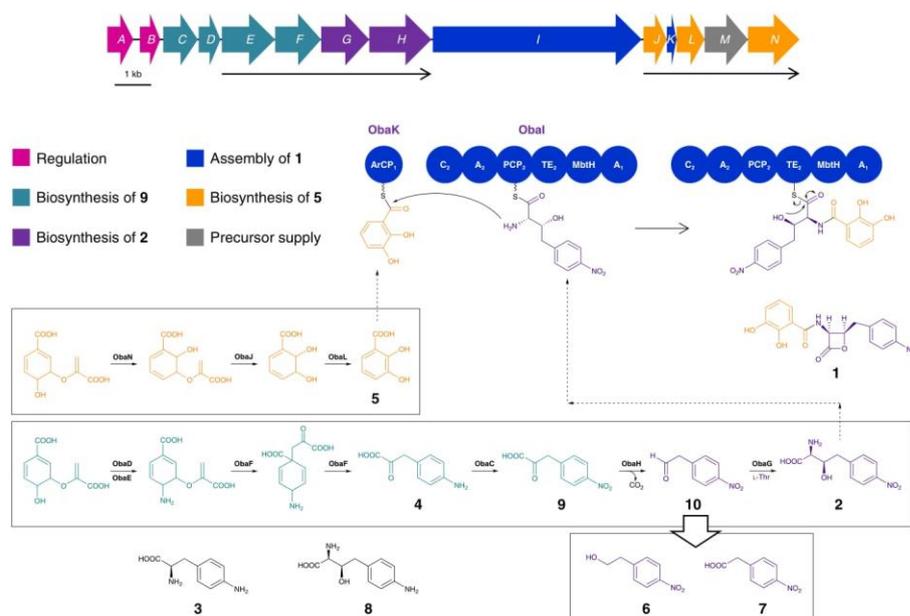


Figure 1 | Gene cluster architecture and proposed biosynthetic pathway for obafluorin biosynthesis in *P. fluorescens* ATCC 39502. Obafluorin protein coding sequences *obaA-N* are labelled and coloured according to related biosynthetic functions, with operons indicated with black arrows beneath the gene cluster. The domain architectures of the aryl carrier protein ObaK and nonribosomal peptide synthetase ObaI are represented (ArCP = Aryl carrier protein; C = Condensation domain; A = Adenylation domain; PCP = Peptidyl carrier protein domain; TE = Thioesterase domain; MbtH = MbtH-like domain), and the anticipated mechanism of TE-catalysed peptide release and β -lactone ring closure is also illustrated. **6** and **7** represent breakdown products of **10**, and were found to accumulate when either the biosynthesis of **5**, or the assembly of **1**, was disrupted. Compounds **3** and **8** were used in chemical feeding experiments in this work, but were not shown to be intermediates in the biosynthesis of **1**.

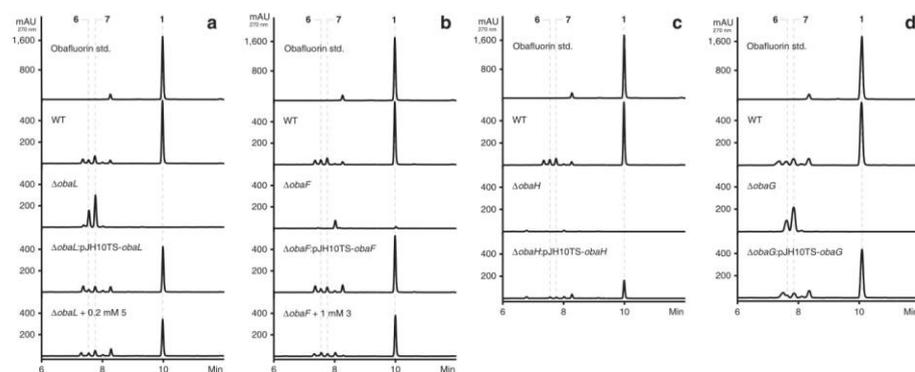


Figure 2 | HPLC profiles for mutagenesis and complementation experiments in selected *oba* genes. (a) *obaL*, (b) *obaF*, (c) *obaH* and (d) *obaG*. Numbered peaks refer to key products and shunt metabolites identified in Fig. 1.

ectopic expression of the deleted gene (to 73% (*obaI*) and 85% (*obaL*) of WT titres). Moreover, by blocking production of the key intermediate 5 (and therefore the pathway endpoint 1), we anticipated that the predicted pathway intermediate 2 (refs 16,17) might accumulate (see Fig. 1). However, after exhaustive LCMS analysis we could not observe any accumulation of 2, nor of the potentially related pathway intermediate (2-amino-3-hydroxy-4-(4-aminophenyl)butanoate (8)). Synthetic standards of both 2 and 8 were prepared to aid this analysis (see Materials and Methods).

The three genes (*obaDEF*) appear sufficient to encode for the production of 4, the biosynthesis of which is well understood from previous genetic studies on chloramphenicol²⁷ and pristinamycin²⁸ biosynthesis. Consistent with these predictions, deletion of the bifunctional gene *obaF* (a bifunctional 4-aminochorismate mutase/4-aminoprephenate dehydrogenase) led to the loss of 1, 6 and 7 production (Fig. 2b). Addition of exogenous 3 was sufficient to restore efficient production of 1 (between 68%–71% of WT titres), as was the ectopic expression of *obaF* (between 73 and 90% of WT titres). We presume that 3 is converted to 4 by transamination *in vivo*. Surprisingly, no accumulation of the pathway intermediate 5 could be observed in the *obaF* mutant.

A putative type II 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHPh) synthase is encoded by *obaM*. Homologues of this gene have been observed in a number of bacterial BGCs^{4,27,29}, where their products are believed to mitigate the aromatic amino acid-based negative feedback of primary metabolic DAHP synthases and drive flux into specialized metabolism. This indicated a likely role for *ObaM* in chorismate supply for the biosynthesis of both 2 and 5. Consistent with this deletion of *obaM* led to the almost complete abolition of 1 biosynthesis (Supplementary Fig. 4), but we were surprised to observe the accumulation of 6 and 7. Moreover, exogenous supply of 5 alone was sufficient to restore production of 1 (68–72% of WT titres), further indicating that biosynthesis of 2 is unperturbed by *obaM* deletion, which appears instead to only impact 5 biosynthesis. Consistent with this, exogenous supply of 3 alone could not recover 1 production but did lead to further elevated levels (five- to sevenfold) of 6 and 7. Ectopic expression of *obaM* recovered 1 production (50–55% of WT titres), precluding the possibility of polar effects on *obaJLN*.

The gene *obaC* encodes a putative non-heme di-iron mono-oxygenase related to AurF from the aureothin pathway (73% query cover, 25% identity)^{30,31} and CmlL from the chloramphenicol pathway (72% query cover, 19% identity)^{32,33}. Both enzymes are arylamine oxygenases which catalyse the conversion of an aromatic amine to a nitro group, and we propose *ObaC* catalyses the conversion of 4 to 4-nitrophenylpyruvate (9). Consistent with this role, deletion of *obaC* abolished production of 1, 6 and 7 (Supplementary Fig. 5). The biosynthesis of all three compounds was re-established after ectopic complementation with *obaC* (81% of WT titres of 1). Surprisingly, only the addition of exogenous synthetic 9 to growing cultures of this mutant was able to restore 1 biosynthesis (to 29% of WT titres), whereas addition of 2 did not. We believe this is likely due to failure of this metabolite to penetrate the *Pseudomonas* cell membrane based on subsequent data below.

At the centre of the BGC is the bimodular NRPS-encoding gene *obaI*, the product of which is predicted to catalyse formation of the amide bond between 2 and 5. The thioesterase (TE) domain of *ObaI* is unusually located between the peptide carrier protein (PCP₂) domain and first adenylation (A₁) domain, and we predict that it catalyses release of the resulting enzyme bound *pseudo*-dipeptide with concomitant formation of the β -lactone moiety to yield 1. Analysis of the A domain sequences of *ObaI* using NRSPredictor2 (ref. 34) identified 5 as the likely substrate of A₁, whereas L-threonine is predicted to be the

substrate of A₂. This means that the domain architecture of *ObaI* (Fig. 1) is further unusual as the putative A₁ domain is located at the C terminus of the enzyme, rather than towards the N terminus as would be expected. Moreover, A₁ is adjacent to an embedded MbtH-like protein domain. MbtH-like proteins are auxiliary proteins required for the activity of many, but not all NRPS A domains^{35–37}. Some A domains require MbtH-like proteins for their functional expression, suggesting they may have a chaperone-like function³⁸. Discrete MbtH-like proteins are the most commonly observed, but the identification of MbtH-like domains embedded within an NRPS, as for *ObaI* here, is rare and has been reported for only two other BGCs, those responsible for the biosynthesis of nikkomycin³⁹ and streptolydigin⁴⁰. While the amino acid residues critical for intra-domain interaction have been well characterized in the latter case, no biochemical data exists to verify that either of these *in cis* domains are functionally essential *in vivo*, and the exact role of MbtH-like proteins in NRPS catalysis is yet to be determined. Deletion of *obaI* abolished production of 1, which was recovered to 79% of WT levels on ectopic complementation (Supplementary Fig. 6). The Δ *obaI* mutant accumulated the shunt metabolites 6 and 7 at similar levels to the WT strain, but, once again, exhaustive analysis of the fermentation extracts did not show any accumulation of 2 or 5, the putative substrates of *ObaI*.

ObaI contains only a single PCP₂ domain where two such domains would be expected for its correct function. Further analysis of the BGC suggested that *obaK* encodes a discrete aryl carrier protein (ArCP₁). The presence of an ArCP is anticipated as this protein function is required for the tethering of 5 (following activation by the A₁ domain) to participate in amide bond formation (Fig. 1). On the basis of these analyses we hypothesize that the unusual positioning of A domains within the assembly line provides access for both to the embedded MbtH-like domain, a likely prerequisite for efficient catalysis. Given the canonical structure and function of NRPS assembly line enzymes, this organization will preclude functional interaction of the A₁ domain and the condensation (C₂) domain, a consequence of which is the requirement for a discrete, and therefore mobile, ArCP₁ (*ObaK*) to enable both interaction with the A₁ domain for acylation with 5 and subsequent interaction with the C₂ domain for peptide bond formation.

Deletion of *obaK* also abolished production of 1, which was recovered by ectopic expression (to 77% of the WT level) (Supplementary Fig. 7). The titres of 6 and 7 also increased significantly (approximately fourfold) but again neither 2 nor 5 was accumulated. Exogenous supply of 5 did not lead to production of 1 confirming that this phenotype is not linked to a lack of 5 biosynthesis. We queried this possibility as *obaK* appears to have evolved by the splitting of an ancestral *entB*-like gene, which usually encode for bidomain proteins involved in the biosynthesis of 5 (refs 25,26). Indeed, the adjacent *obaJ* gene corresponds to the N-terminal isochorismatase domain of EntB and is required for production of 5 during biosynthesis of 1. The biosynthesis of 5 does not require the ArCP domain of EntB⁴¹ consistent with our various results and biosynthetic hypothesis.

Our analysis thus left the biosynthesis of 2 as the remaining unknown, and only the products of *obaG* and *obaH* to be functionally assigned. *obaG* encodes a putative SHMT/L-TA while *obaH* encodes a putative ThDP-dependent pPDC. Mutation of both genes abolished the production of 1, but the accumulation of 6 and 7 was only observed in the *obaG* mutant (Fig. 2c,d). Again, exogenous addition of 2 did not recover 1 production in these mutants whereas ectopic expression of the deleted genes did (to 62% (*obaG*) and 29% (*obaH*) of WT titres). These mutational data are consistent with the role of *ObaG* as the enzyme directly responsible for 2 production as proposed in Fig. 1, but did not

provide direct support for **2** as a biosynthetic intermediate. In contrast, the accumulation pattern for the shunt metabolites **6** and **7** strongly suggested that 4-nitrophenylacetaldehyde (**10**) might be a key intermediate. This reactive and potentially toxic molecule is likely to be catabolized to the shunt metabolites **6** and **7** via standard detoxification pathways before accumulating to any degree. Consistent with this hypothesis, addition of exogenous **10** to growing cultures of the $\Delta obaG$ mutant did not restore **1** production but did lead to elevated levels of **6** and **7**. We also noted that chemical complementation of mutants deficient in **5** biosynthesis are sensitive to the concentration of exogenous material added. Production of **1** was re-established when **5** was added to give a final concentration of 0.2 mM, but at higher concentrations (for example, ≥ 0.5 mM) both growth of *P. fluorescens* and biosynthesis of **1** were significantly affected. As noted above we could not observe the presence of **2** or **5** in any of the mutants anticipated to accumulate them. Additionally, in mutants where **2** or its aldehyde precursor **10** might have accumulated, we always observed elevated levels of the shunt metabolites **6** and **7**, breakdown products of aldehyde **10**, a highly reactive and toxic molecule. Taken together these combined data suggest that the biosynthetic pathway is tightly regulated to avoid the accumulation of intermediates deleterious to cell growth and survival. Interestingly, the intermediary of free phenylacetaldehyde, a reactive aldehyde structurally similar to **10**, is avoided in the recently reported ripostatin biosynthetic pathway via the activity of a pyruvate dehydrogenase-like complex which functions to decarboxylate phenylpyruvate and then retain the phenylacetaldehyde product as a phenylacetyl-S-carrier protein species⁴². Finally, bioinformatics suggests the obafluorin BGC is regulated by quorum sensing under the control of the *luxIR* homologues *obaAB*.

2 is assembled by sequential activity of ObaH and ObaG. The essential nature of *obaG* and *obaH*, in conjunction with the accumulation pattern of **6** and **7** by their mutants, led us to hypothesize that the biosynthesis of **2** involves ThDP-dependent decarboxylation of **9** by ObaH to yield the aldehyde **10**. This would then be used as a substrate by the pyridoxal-5'-phosphate (PLP) dependant enzyme ObaG acting as an aldolase to give **2**. Both enzymes were readily expressed as soluble proteins in hexahistidine-tagged form using *E. coli* NiCo21(DE3) pLysS. They were purified using standard procedures and their identity verified by tandem MS.

Due to the original proposal of Herbert and Knaggs that a pPDC-like enzyme might produce an acyloin intermediate (Supplementary Fig. 1)¹⁶ we first tested ObaH with **9** plus glyoxylate as substrates using a discontinuous format coupled to independent LCMS/MS and HPLC-UV assays. In none of these reactions could we detect any acyloin products whereas the ketoacid substrate was depleted and aldehyde **10** accumulated whenever ObaH was incubated with **9**, clearly identifying it as a **9** decarboxylase (Fig. 3a). To examine the ability of the putative *l*-TA-like enzyme ObaG to produce **2**, it was incubated with varying concentrations of **10** and glycine using the same assay format as for ObaH. Under none of these conditions could we observe production of **2**. At this point we considered the possibility that ObaG might actually be a transaldolase and incubated it with **10** and either *l*-serine or *l*-threonine instead of glycine, and were gratified when we observed excellent production of the anticipated intermediate **2** only when *l*-threonine was present in the reaction mixture (Fig. 3b). Moreover, when [U - $^{13}C_4$, $^{15}N_1$]*l*-threonine was used as substrate we were able to demonstrate, by ^{13}C NMR spectroscopy and high-resolution (HR)-LCMS/MS, the regioselective transfer

of three heavy isotope atoms to the appropriate positions in the product **2**, in addition to the formation of [$1,2$ - $^{13}C_2$]acetaldehyde (Figs 3c and 4). The structure and stereochemistry of isolated **2** was then verified by comparison of its NMR spectra and optical rotation to those of previously synthesized material⁴³. To confirm the biosynthetic relevance of *l*-threonine *in vivo* we fed [U - $^{13}C_4$, $^{15}N_1$]*l*-threonine to growing cultures of the WT producer and examined the resulting **1** by HR-LCMS/MS analysis. This indicated the highly efficient, site specific incorporation of a $^{13}C_2^{15}N_1$ unit as anticipated (Supplementary Fig. 8). In a final experiment, we were able to couple the ObaH decarboxylase and ObaG *l*-threonine transaldolase (*l*-TTA) reactions *in vitro* (Fig. 3c,d), to catalyse the formation of **2** using *l*-threonine and **9** as the only substrates. The proposed mechanism for the biosynthesis of **2** by ObaG is shown in Supplementary Fig. 9.

Having characterized the substrates and function of both ObaH and ObaG, we sought to further characterize ObaG by confirming its identity as a PLP-dependent enzyme and attempted to collect preliminary kinetic data. Despite being identified as a putative SHMT, ObaG shares relatively little sequence similarity to *bona fide* bacterial SHMTs (25% amino acid sequence identity), but does comprise several residues crucial for PLP-dependent activity, most importantly the lysine residue required for internal aldimine formation (Supplementary Fig. 10). ObaG displays a UV/Vis spectrum (Supplementary Figs 11a and b) characteristic of PLP-dependent proteins⁴⁴, comprising absorption maxima at 340 and 390 nm that correspond to the equilibrium established between the enolimine and ketoenamine forms of PLP respectively. ObaG is unusual in that it exhibits a distinct salmon pink colour in solution (Supplementary Fig. 11c), which creates an additional absorbance maximum at 512 nm (Supplementary Figs 11a and b). Treatment of ObaG in the absence of excess PLP with 10 mM *l*-penicillamine led to the anticipated formation of a thiazolidine adduct (Supplementary Fig. 11d) determined by UV absorbance at 340 nm (Supplementary Fig. 11a), with concomitant loss of the ketoenamine peak at 390 nm, as observed previously by Lowther *et al.*⁴⁴ for serine palmitoyltransferase. ObaG was further incubated with NaBH₄ (1 mM) resulting in the formation of a peak at 330 nm consistent with the formation of the reduced ObaG-PLP amine adduct⁴⁵ (Supplementary Figs 11b and e). In both experiments a decrease in the 512 nm peak was also observed. Taken together these results strongly support ObaG being a PLP-dependent enzyme.

Having characterized the *l*-TTA reaction catalysed by ObaG, and its PLP-dependence, we acquired preliminary kinetic data. A time course was determined using standard *l*-TTA assay conditions showing that equilibrium between substrate **10** and product **2** was achieved within ~1 h after reaction initiation (Supplementary Fig. 12a). Unfortunately, the UV/Vis spectrum of **10** precluded the use of a continuous coupled assay based on acetaldehyde accumulation due to overlap with that of the measured side-product, and its inherent reactivity confounded discontinuous methods measuring product formation with varying **10**. The best results were achieved by varying *l*-threonine concentration using a discontinuous HPLC-based approach. Single-substrate kinetic analysis of His₆-ObaG revealed typical Michaelis-Menten kinetics with respect to varying *l*-threonine, yielding kinetic constants of $K_m = 40.2 \pm 3.8$ mM and $k_{cat} = 62.9 \pm 1.9$ min⁻¹ (Supplementary Fig. 12b).

Preliminary investigation of the NRPS ObaI. In a final biochemical experiment we were able to express the NRPS ObaI as the full length (211.5 kDa) protein in its *apo*-form and probe the ability of its two A domains to activate various amino

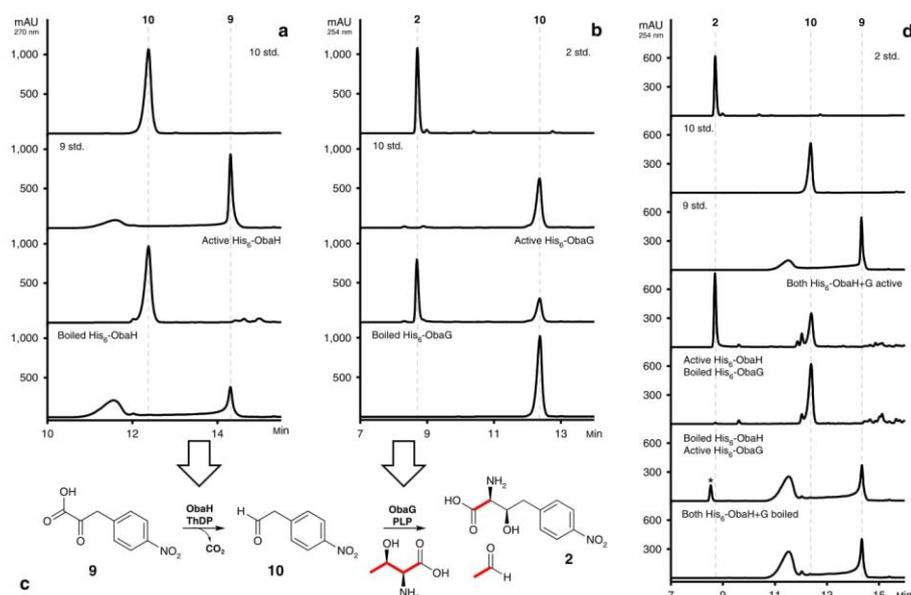


Figure 3 | Experimental characterization of the biosynthesis of 2. (a) His₆-ObaH decarboxylase activity assay HPLC profiles. (b) His₆-ObaG L-TTA activity assay HPLC profiles. (c) Illustration of the biosynthesis of 2 in a two-step reaction catalysed by consecutive action of ObaG and H. The His₆-ObaG L-TTA activity assay was also performed using [¹³C₄,¹⁵N₁] L-threonine, and terminated reactions were purified and analysed by NMR (Fig. 4) to show ¹³C incorporation into 2. The result of this experiment is also illustrated here. (d) Coupled decarboxylase and L-TTA assay HPLC profiles. 9 was also found to be a poor substrate for the ObaG-catalysed L-TTA reaction, yielding unknown product*. In each case substrate/product peaks are numbered in accordance with Fig. 1.

acids using a discontinuous hydroxamate formation assay⁴⁶ (Methods section). While this approach did not allow us to dissect the activities of the two A domains independently, it did suggest that 2 is the preferred substrate for the A₂ domain, but that it displays relaxed substrate specificity with activation of L-tryptophan and L-serine being observed (Supplementary Fig. 13). It should be noted that NRPSpredictor2 (ref. 34) identified L-threonine as the substrate for the A₂ domain. The A₁ domain, which is predicted to activate 5, also appears to have relaxed specificity as benzoic acid was activated in addition to 5. Despite the apparent flexibility of both A domains we could not observe the production of obafluorin congeners in the WT strain, although such molecules might be sensitive to proteolytic degradation. Most likely, and consistent with the observations of others, while the A domains of intact ObaI show an inherent substrate flexibility, the C₂ domain is likely to exert a gatekeeper function with regards to the selection of substrates for carrier protein acylation and subsequent amide bond formation⁴⁷.

Phylogenetic analysis of L-TTAs. L-TTA activity like that of ObaG here has previously been described only in the biosynthesis of 4-fluoro-L-threonine⁴⁸ and in 5'-C-glycyluridine (GlyU), a key constituent of lipopeptidyl nucleoside natural products⁴⁹ (Supplementary Fig. 14). Both 4-fluoro-L-threonine transaldolase

(FTase) and L-threonine:uridine-5' transaldolases of the LipK family were all identified as putative SHMTs, but have been shown to catalyse L-threonine β-substitutions via the sequential breaking and forming of Cα–Cβ bonds. We have now undertaken detailed phylogenetic analyses of PLP-dependent SHMTs, L-TAs and the L-TTAs described here (Fig. 5)⁵⁰. This reveals that bacterial L-TTAs and SHMTs diverged from each other to form discrete clades, and that they and the TAs diverged from an earlier common ancestor. It is particularly notable that the common ancestor appears to have diverged into enzymes that all utilize L-threonine, other than the SHMTs. While the evolutionary relationship between ObaG and FTase is not so easily determined when comparing different phylogenies (Supplementary Figs 15 and 16), what is clear is that they are distinct from the L-threonine:uridine-5' transaldolase (LipK) clade. In addition, OrfA—identified as a glycine hydroxymethyltransferase and proposed as a biosynthetic enzyme required for the biosynthesis of alanylclavam in the cephamycin/clavulanic acid/5S-clavam producer *Streptomyces clavuligeri*⁵¹—also consistently grouped within the L-TTA clade. The biochemical characterization of this enzyme has not yet been reported, but based on our analysis we suspect that it has been incorrectly annotated and is in fact an L-TTA using L-threonine as a substrate rather than an L-TA using glycine as previously proposed (Supplementary Fig. 14).

Discussion

On the basis of our bioinformatics analysis combined with mutational, biochemical and chemical feeding results we propose the biosynthetic pathway for **1** shown in Fig. 1. The pathway

involves a unique, non-canonical NRPS which carries an integrated MbtH-like protein domain and contains a central TE domain which is likely to be responsible for β -lactone formation. Central to the pathway, however, is a new route for

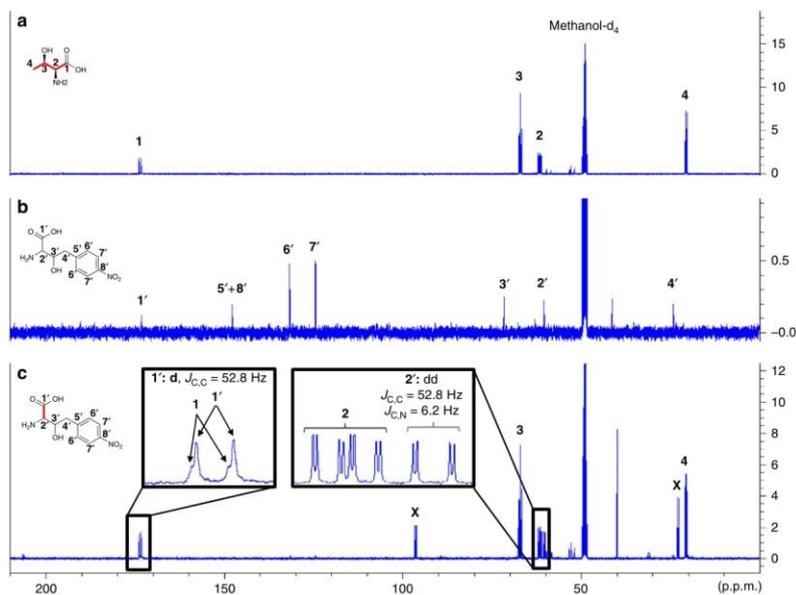


Figure 4 | Characterization of the biosynthesis of **2 by ObaG using ^{13}C NMR.** (a) ^{13}C NMR spectrum of $[\text{U-}^{13}\text{C}_4,^{15}\text{N}]$ L-threonine in methanol- d_4 . (b) ^{13}C NMR spectrum of the synthetic reference **2**. (c) ^{13}C NMR spectrum of enzymatically synthesized **10**. Signals marked with X correspond to the two acetal forms of acetaldehyde and glycerol, being formed after release of acetaldehyde and therefore proving threonine L-TTA activity of ObaG.

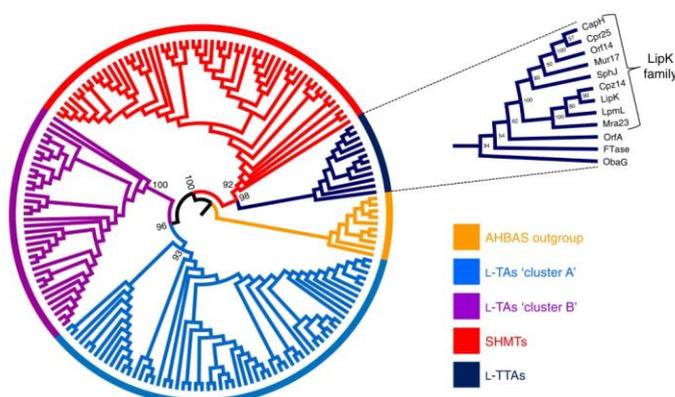


Figure 5 | Maximum likelihood tree of L-TAs, SHMTs and L-TTAs. A set of 3-amino-5-hydroxybenzoate synthase (AHBAS) amino acid sequences serve as the outgroup. RAxML likelihood values at the root nodes for SHMT and L-TA cluster clades are annotated. The clade comprising the L-TTA ObaG characterized in this work has been expanded and annotated with RAxML likelihood values and the identity of amino acid sequences represented.

biosynthesis of the homologated *L*-threo- β -hydroxy- α -amino acid **2** via sequential decarboxylation and transaldolase reactions. *L*-TTA reactions are rare and have been reported previously only for the biosynthesis of 4-fluoro-*L*-threonine and GlyU from fluoroacetaldehyde⁴⁸ and uridine-5'-aldehyde⁴⁹, respectively (Supplementary Fig. 14), both of which exhibit the *L*-threo stereochemistry of **2**. Detailed phylogenetic analysis has shed light on the evolutionary origins of the *L*-TA (which use glycine as substrate) and *L*-TTA families (Fig. 5), and suggests that the two families diverged from a common ancestor with the transaldolase family subsequently splitting again into lineages that specifically utilize either *L*-threonine (*L*-TTAs) or *L*-serine (SHMTs) as substrates. With this knowledge in hand we used a phylogeny-based, mechanism-guided genome mining approach to identify BGCs containing *L*-TTAs, and therefore biosynthetic pathways likely involving a β -hydroxy- α -amino acid intermediate. This led to the identification of a putative *L*-TTA involved in the biosynthesis of a 5S-clavam metabolite, indicating phylogenetic analysis as a potentially general genome mining approach for the identification of further *L*-TTAs.

We also identified multiple BGCs likely to encode for the production of **1** or very closely related compounds in the genomes of several *Pseudomonas* environmental isolates⁵², *Burkholderia* strains (Genome sequences for *Burkholderia* species containing *oba* like BGCs: *Burkholderia stagnalis* (NZ_LPGD01000044.1); *Burkholderia diffusa* (NZ_LOTC01000033.1), *Burkholderia territorii* (NZ_LOSY01000044.1), *Burkholderia ubonensis* (NZ_CP013463.1.)) and a *Chitiniphilus* strain (Genome sequences for *Chitiniphilus* species containing *oba* like BGCs (GenBank Accession: NZ_KB895358.1)). The relatively common occurrence of the **1** BGC is consistent with the original report describing its discovery which found that numerous *Pseudomonas* isolates, from a range of geographical locations, were able to produce **1**¹³. In several of the sequenced genomes, but not all, a set of genes encoding a glycine cleavage system⁵³ is located adjacent to the putative biosynthetic genes. We also identified an apparently intact glycine cleavage system (Supplementary Table 1) adjacent to the BGC for **1**.

The utility of *L*-TAs for the enzymatic synthesis of *L*-threo- β -hydroxy- α -amino acids from an aldehyde and glycine has been extensively investigated, but is limited due to low synthetic yields and modest diastereoselectivity^{10,11}. The low yields are in part due to the reversible nature of the reaction (retroaldol cleavage of *L*-threo- β -hydroxy- α -amino acids) and an equilibrium which favours aldehyde and glycine. Due to this synthetic reactions are commonly run in the presence of an excess of glycine to shift the equilibrium to the aldol side. The highly efficient reversible nature of this reaction has, however, been harnessed for the resolution of racemic mixtures to provide *D*-threo-isomers⁵⁴.

In contrast *ObaG* gives **2** as a single stereoisomer in an excellent 55–59% yield based on accumulation of **2** (Supplementary Fig. 12a) without the need for an excess of *L*-threonine. On the basis of this result we undertook preliminary investigations into the ability of *ObaG* to accept the alternative aldehyde substrates phenylacetaldehyde and benzaldehyde (Supplementary Figs 17a and b). Incubation with phenylacetaldehyde led to a single product in excellent yield (~45%, based on consumption of starting material) whose LCMS profile was consistent with the expected product (2*S*)-amino-(3*R*)-hydroxy-4-phenylbutanoate. The reaction with benzaldehyde was less efficient leading to two products in a 1:2 ratio and poor overall yield (<20%). The LCMS profile of these was consistent with the production of both *L*-threo and *L*-erythro-phenylserine, respectively (an authentic standard of the *threo* diastereoisomer was utilized). These data suggest that *L*-TTAs may offer an alternative to *L*-TAs for the synthesis of *L*-threo- β -hydroxy- α -amino acids. The equilibrium appears to lie in the favour of the aldol reaction and might be improved further

through the use of excess *L*-threonine and the addition of glycerol which reacts preferentially with acetaldehyde, shifting the equilibrium further in the direction of the target molecule. This is consistent with our observation that *ObaG* is a poor catalyst for the reverse transaldol reaction when incubated with acetaldehyde and racemic **2** made synthetically (Supplementary Fig. 17c). Moreover, our preliminary data suggest that there is a potential for high diastereoselectivity depending on the substrate aldehyde which is utilized. Future work in our lab will include additional biochemical and kinetic analysis of *ObaG*, and a more comprehensive investigation of its potential utility for the synthesis of enantiomerically pure *L*-threo- β -hydroxy- α -amino acids.

In summary, we have delineated the entire biosynthetic pathway to **1**, and identified a distinct enzyme family for the synthesis of β -hydroxy- α -amino acids, privileged chiral building blocks in a variety of pharmacologically and agriculturally important natural products and medicines. *L*-TTAs perform a unique transformation, incorporating two stereocentres in one enzymatic step, with high diastereoselectivity and excellent yields. Thorough phylogenetic investigation revealed that the *L*-TTAs form their own discrete evolutionary lineage, distinct from *L*-TAs and SHMTs, shedding light on their own evolutionary history.

Materials

General. Reagents and chemicals were purchased from Alfa-Aesar, Sigma-Aldrich, Santa Cruz Biotechnology, Inc., Amatek Chemical Co., Ltd., and BD Biosciences, and were used without further purification. All strains used in this study are listed in Supplementary Table 2. *Pseudomonas fluorescens* ATCC 39502 was purchased from the American Type Culture Collection (ATCC). All strains were maintained on solid Luria-Bertani medium (with appropriate selection) at 37 °C (*E. coli*) or 28 °C (*P. fluorescens*). All solvents used for HPLC were obtained from Fisher Scientific at least of HPLC grade and were filtered before use.

Instrumentation. NMR spectra were recorded on a Bruker AVANCE III 400 MHz spectrometer. Chemical shifts are reported in parts per million (p.p.m.) relative to the solvent residual peak of chloroform-*d*₁ (¹H: 7.24 p.p.m., singlet; ¹³C: 77.00 p.p.m., triplet), methanol-*d*₄ (¹H: 3.30 p.p.m., quintet; ¹³C: 49.00 p.p.m., septet) or water-*d*₂ (1H: 4.79 p.p.m.). Semi-preparative and analytical HPLC was performed using an 1100 system (Agilent Technologies). Extracted samples from mutagenesis and complementation experiments were analysed using a Gemini 3 μ m NX-C18 110 Å, 150 \times 4.6 mm column (Phenomenex) with a gradient elution: MeCN/0.1% (v/v) TFA (H₂O) gradient from 10/90 to 100/0 0–15 min, 100/0 for 15–16 min, gradient to 10/90 16–16.50 min and 10/90 for 16.50–23 min; flow rate 1 ml min⁻¹; injection volume 10 μ l. Biochemical assays were analysed using a Synergi 4 μ m Fusion-RP 80 Å LC column 250 \times 10 mm with a gradient elution: MeOH/0.1% (v/v) TFA (H₂O) gradient from 10/90 to 100/0 0–14 min, 100/0 for 14–18 min, gradient to 10/90 18–18.50 min and 10/90 for 18.50–23 min; flow rate 1 ml min⁻¹; injection volume 5 μ l (15 μ l for the reverse reaction experiments). UPLC-MS measurements were performed on a Nexera X2 liquid chromatograph (LC-30AD) LCMS system (Shimadzu) connected to an autosampler (SIL-30AC), a Prominence column oven (CTO-20AC) and a Prominence photo diode array detector (SPD-M20A). A Kinetex 2.6 μ m C18 100 Å, 100 \times 2.1 mm column (Phenomenex) was used for LCMS. The UPLC-System was connected with a LCMS-IT-TOF Liquid Chromatograph mass spectrometer (Shimadzu). Solid phase extraction was carried out using Discovery DSC-18 SPE Tubes, filled with 1,000 mg of octadecyl-modified, endcapped silica gel (Supelco). The specific optical rotation of compounds was measured with a Model 341 Polarimeter (PerkinElmer, Inc.).

Cloning. All primers used in this study are reported in Supplementary Table 3. Δ *oba* strains were generated using the suicide vector pTS1, constructed as part of this work (Supplementary Note 1). Primers were designed to amplify 800–1,200 bp flanking regions of a selected *oba* protein coding sequences (PCSs) for cloning into pTS1 between *Xba*I and *Avr*II sites, and *Avr*II and *Bmt*I sites. Flanking regions were designed to comprise 10–50 PCS codons at either end of the *oba* gene to be deleted to minimize polar effects, leaving a truncated chromosomal copy of the gene with an in-frame deletion and internal *Avr*II site cloning artefact following double homologous recombination. For genetic complementation, WT copies of *oba* PCSs were cloned from start to stop as *Bmt*I-*Kpn*I/*Xba*I fragments into pJH10TS (Supplementary Note 2) for introduction and ectopic expression in the relevant mutant strains. For protein purification and expression, WT PCSs for *obaG-I* were cloned as *Nde*I-*Xho*I fragments into pET28a(+) for expression in *E. coli* NiCo21(DE3) (NEB) pLysS.

Mutagenesis and complementation experiments. pT51 knockout constructs were introduced into *P. fluorescens* ATCC 39502 via conjugation from *E. coli* S17-1 λ pir and single-crossover mutants were selected for on LB Tc²⁵. Positive colonies were cultured overnight in antibiotic-free medium to allow time for a second cross-over event to occur. *sacB* counter-selection could then be performed by plating culture dilutions on 10% sucrose to select against retention of the pT51 vector backbone. Colony PCR was then performed to identify double cross-over mutants from WT colonies, and Sanger sequencing (Eurofins Genomics) was performed to confirm expected deletions. pHHOTS complementation constructs were also conjugated into the relevant *Oba* strains and positive clones were selected for on LB Tc²⁵. Clones were screened by colony PCR and confirmed by sequencing. *Oba* strains were complemented chemically where possible by introduction of compounds when production cultures were established. Compounds fed and the concentrations used are described in the main text.

Analysis of metabolite production. WT and recombinant *P. fluorescens* ATCC 39502 strains were grown in Obafluorin Production Medium (OPM) (Yeast extract 0.5%, D-glucose 0.5%, MgSO₄ × 7H₂O 0.1%, and FeSO₄ 0.1%, dissolved in Milli-Q (Merck Millipore) filtered water). A toothpick was used to inoculate 100 ml of OPM seed culture (250 ml Erlenmeyer flask) from a 40% glycerol stock (stored at -80 °C), with subsequent growth for 24 h at 25 °C, 300 r.p.m. 1 ml of this culture was used to inoculate a 100 ml (500 ml flask) OPM production culture which was incubated under the same conditions for 14 h. Samples were prepared for HPLC/LCMS analyses by extracting 1 ml of culture broth with an equal volume of ethyl acetate by mixing at 1,400 r.p.m. for 15 min. Samples were then centrifuged (1,616g for 15 min), and the organic phase was collected and evaporated. The resulting extract was dissolved in MeCN (500 μ l) and centrifuged (1,616g for 20 min) to remove any remaining cell debris.

Protein expression and purification. *E. coli* NiCo21(DE3) (NEB) pLysS strains carrying pET28a(+) *obaG*, pET28a(+) *obaH* and pET28a(+) *obaI* were cultivated in Terrific Broth (TB) at 28 °C and 250 r.p.m. on a rotary shaker until A_{600 nm} = 0.5. Protein expression was induced by addition of 0.1 mM IPTG and incubation continued at 18 °C and 200 r.p.m. for 18 h. Cells were pelleted at 2,415g and 4 °C, and were subsequently re-suspended in buffer containing 25 mM HEPES (ObaG and ObaH)/50 mM Tris-HCl (ObaI) at pH 7.8 containing NaCl (300 mM), MgCl₂ (15 mM) and glycerol (10%). ObaG and ObaH buffers were further supplemented with PLP (0.4 mM) and ThDP (0.5 mM) respectively. After disruption with an EmulsiFlex-B15 high pressure homogeniser (Avestin, Inc.), cells were pelleted at 26,892g and 4 °C for 30 min. The lysed supernatant was incubated with chitin resin with gentle mixing for 30 min to remove any endogenous *E. coli* metal binding proteins. Eluted sample was loaded onto a HisTrap excel (GE Healthcare) Ni-NTA column using an AKTA pure (GE Healthcare) system. Proteins were washed in 5 CV of their respective buffers containing 10, 20, 30 and 50 mM (His₆-ObaG and His₆-ObaH only) imidazole concentrations. His₆-ObaI was eluted with 20 CV of 50 mM imidazole, and His₆-ObaG and His₆-ObaH were eluted in 20 CV of 250 mM imidazole, all in 2 ml fractions. Oba protein-containing fractions were combined and applied to Amincon columns (30 kDa MWCO) and diluted > 1,000 × to remove imidazole, before being concentrated. The His₆-ObaI sample was further purified by size exclusion over a HiLoad 16/600 Superdex 200 pg column (GE Healthcare). Eluted His₆-ObaI-containing fractions were combined and concentrated in an Amincon column (30 kDa MWCO). Protein samples were stored at 4 °C for *in vitro* assays and long-term storage was at -80 °C in their respective buffers supplemented with 20% glycerol. Recombinant protein samples were run on SDS-PAGE gels to confirm their size (Supplementary Fig. 18). Protein bands were subsequently cut out, washed, reduced, alkylated and treated with trypsin according to standard procedures adapted from Shevchenko *et al.*⁵⁵ The tryptic peptide fragments were analysed by mass spectrometry to further confirm protein identity, using an autoflex Speed MALDI-TOF/TOF mass spectrometer (Bruker Daltonics GmbH).

***In vitro* His₆-ObaG discontinuous L-TTA activity assay.** Reactions were performed in 100 μ l reaction volumes comprising 50 μ M His₆-ObaG, 10 mM glycine/L-serine/L-threonine or [U-¹³C₄,¹⁵N]-L-threonine (98% isotopic purity) (for subsequent NMR experiments), and 10 mM **10**, all in His₆-ObaG buffer (25 mM HEPES pH 7.8, 300 mM NaCl, 15 mM MgCl₂, 0.4 mM PLP). Reactions were initiated by introduction of the enzyme and were incubated at 27 °C and 700 r.p.m. for 2 h. A boiled enzyme sample was used as a negative control. Reactions were terminated by addition of MeOH (100 μ l) and were incubated at -20 °C for 1 h to ensure full enzyme precipitation. Precipitated enzyme was pelleted at 1,616g for 30 min before analysis by HPLC. Time course data were similarly collected by terminating the reaction at a range of time points up to 5 h and all time points were assayed in triplicate.

The amenability of His₆-ObaG to alternative substrates was explored by performing the assay described above but using 10 mM of either benzaldehyde or phenylacetaldehyde instead of **10** as a co-substrate with L-threonine. The reverse reaction to generate **10** and L-threonine using **2** with and without acetaldehyde as substrates with His₆-ObaG was also performed as above but using 20 mM concentrations of starting substrates.

Single-substrate kinetic analysis was carried out by performing the L-TTA activity assay with varying concentrations of L-threonine (1–200 mM). Reactions were performed using 25 μ M enzyme and were incubated at 27 °C, 700 r.p.m. for 4 min, before quenching, sample processing and HPLC analysis as described previously. Five replicates were performed for each concentration of L-threonine assayed, and data were fitted to the Michaelis-Menten equation using GraphPad Prism 5.04 (GraphPad Software, Inc., La Jolla, USA).

***In vitro* His₆-ObaH decarboxylase activity assay.** Reactions were performed in 100 μ l reaction volumes comprising 50 μ M His₆-ObaH and 10 mM **9** or phenylpyruvate, all in His₆-ObaH buffer (25 mM HEPES pH 7.8, NaCl (300 mM), MgCl₂ (15 mM) and ThDP (0.5 mM)). Reactions were initiated by introduction of the enzyme and were incubated at 27 °C and 700 r.p.m. for 5 min. A boiled enzyme sample was used as a negative control. Reactions were terminated as described previously. HPLC and LCMS analysis was performed as described for the His₆-ObaG L-TTA activity assay.

Coupling of His₆-ObaH and His₆-ObaG reactions. His₆-ObaG was first exchanged into His₆-ObaH buffer (see previous) to avoid reaction of **9** with unbound PLP. Reactions were performed in 100 μ l reaction volumes comprising 50 μ M His₆-ObaG, 50 μ M His₆-ObaH, 10 mM **9** and 10 mM L-threonine, all in His₆-ObaH buffer. Reactions were initiated by introduction of the enzyme and were incubated at 27 °C and 700 r.p.m. for 2 h. Control reactions were also performed in which one or both of His₆-ObaG and His₆-ObaH were boiled before the reaction. Reactions were terminated by addition of MeOH (100 μ l) and were incubated at -20 °C for 1 h to ensure full enzyme precipitation. Precipitated enzyme was pelleted at 1,616g for 30 min before analysis by HPLC.

PLP-dependence of His₆-ObaG. Recombinant His₆-ObaG was incubated with L-penicillamine⁴⁴. Excess PLP was removed from His₆-ObaG samples by exchanging into a reaction buffer comprising 25 mM HEPES pH 7.8, 300 mM NaCl, and 15 mM MgCl₂, using an Amincon column (30 kDa MWCO). Reactions were performed in 600 μ l reaction volumes comprising 20 μ M His₆-ObaG and were initiated by addition of L-penicillamine (dissolved in reaction buffer) to a final concentration of 10 mM. Ultraviolet-visible spectra were recorded over 30 min using a Lambda 35 UV/Visible Spectrophotometer (PerkinElmer).

A second spectrophotometric assay was performed using His₆-ObaG samples with excess PLP removed in which protein was treated with NaBH₄ to reduce the His₆-ObaG-PLP aldimine to form an amine adduct⁴⁵. Reactions were performed on ice for 15 min in 600 μ l volumes comprising 20 μ M His₆-ObaG and were initiated by addition of NaBH₄ (dissolved in reaction buffer) to a final concentration of 1 mM. Samples were analysed by UV/Visible spectrophotometry.

His₆-ObaI hydroxylamine-trapping assay. Reaction mixtures comprised 8.5 μ M His₆-ObaI, 50 mM Tris-HCl pH 8, ATP (2.25 mM), hydroxylamine (150 mM), amino acid substrate (5 mM), MgCl₂ (15 mM) in a final volume of 300 μ l, and were allowed to proceed at 28 °C for 5 h. Boiled enzyme and no substrate reactions were performed as negative controls. Following reaction termination by addition of quenching solution (10% FeCl₃ × 6H₂O and 3.3% trichloroacetic acid made up in 0.7 M HCl), and centrifugation to pellet precipitated protein, samples were transferred to cuvettes and were measured at A_{540 nm} on a Spectronic Biomate 3 (Thermo Fisher Scientific).

Chemical synthesis and isolation of substrates. For the isolation of **1** and its methanolysis product, the supernatant from 61 of culture broth was subjected to the liquid-liquid partition using equal volume of ethyl acetate. The organic fraction was concentrated under reduced pressure and the residue dissolved in MeCN. This was subjected to preparative reversed-phase HPLC (C18, 150 × 21.2 mm, 110 Å. Phenomenex; A: water; B: MeCN; gradient 0–5 min 10% B (v/v), 5–35 min, 10–100% B (v/v), 35–40 min 100% B (v/v), 40–41 min 100–10% B (v/v), 41–45 min, 10% B (v/v); flow rate was 20 ml min⁻¹; monitored at 250 nm). The fractions containing **1** and its methanolysis product were combined and further purified by semi-preparative reversed phase HPLC (C18, 150 × 10 mm, 110 Å. Phenomenex; A: water; B: MeCN; isocratic 50% B). **1**: Pale yellow solid, (13 mg). ¹H NMR (400 MHz, MeCN-d₃) δ 8.23 (1H, d, *J* = 8.8 Hz), 8.10 (2H, d, *J* = 8.8 Hz), 7.45 (2H, d, *J* = 8.8 Hz), 7.20 (1H, dd, *J*₁ = 8.2 Hz, *J*₂ = 1.27 Hz), 7.04 (1H, dd, *J*₁ = 7.9 Hz, *J*₂ = 1.2 Hz), 6.82 (1H, t, *J*₁ = 7.96 Hz, *J*₂ = 5.1 Hz), 5.75 (1H, dd, *J*₁ = 8.5 Hz, *J*₂ = 6.2 Hz), 5.05 (1H, m), 3.38 (1H, dd, *J*₁ = 15.1 Hz, *J*₂ = 5.1 Hz), 3.21 (1H, dd, *J*₁ = 15.1 Hz, *J*₂ = 5.1 Hz); ¹³C NMR (100 MHz, MeCN-d₃) δ 171.47, 169.11, 150.43, 148.08, 147.02, 145.41, 131.20, 124.64, 120.50, 120.07, 118.68, 114.86, 78.50, 59.87, 36.16 p.p.m. HRMS (*m/z*): [M + H]⁺ calculated for C₁₇H₁₄N₂O₂, 359.0874, found, 359.0872. The NMR data is consistent with published values⁵⁶. Methanolysis product: Pale yellow solid (23 mg). ¹H NMR (400 MHz, MeCN-d₃) δ 8.12 (1H, d, *J* = 8.8 Hz), 7.60 (1H, d, *J* = 8.6 Hz), 7.48 (2H, d, *J* = 8.7 Hz), 7.27 (1H, d, *J* = 8.4 Hz), 7.02 (1H, d, *J* = 7.7 Hz), 6.80 (1H, d, *J* = 8.1 Hz), 4.77 (1H, dd, *J*₁ = 2.3 Hz, *J*₂ = 8.9 Hz), 4.46 (1H, m), 3.72 (3H, s), 3.00 (1H, dd, *J*₁ = 14 Hz, *J*₂ = 5.2 Hz), 2.93 (1H, dd, *J*₁ = 14 Hz, *J*₂ = 5.2 Hz); ¹³C NMR (100 MHz, MeCN-d₃) δ 171.90, 171.56, 150.51, 148.13, 147.78, 147.17, 131.90, 124.65, 120.34, 120.12, 119.04, 115.69, 72.91, 57.64, 53.58, 41.08 p.p.m. HRMS (*m/z*): calculated for [M + H]⁺ C₁₈H₁₆N₂O₂, 391.1136, found, 391.1139.

Ethyl 5-(4-nitrobenzyl)-4,5-dihydrooxazole-4-carboxylate (**11**) was synthesized according to a literature procedure⁵⁶ using *p*-nitrophenylacetaldehyde (990 mg, 6 mmol) and ethyl isocyanate (750 mg, 6.6 mmol) to yield 910 mg of the racemic *trans*-diastereoisomer. Yield: 55%. ¹H NMR (400 MHz, CDCl₃) δ 8.08 (2H, d, *J* = 8.8 Hz), 7.36 (2H, d, *J* = 8.8 Hz), 6.87 (1H, d, *J* = 9.4 Hz), 4.72 (1H, dd, *J*₁ = 9.4 Hz, *J*₂ = 2.0 Hz), 4.41–4.36 (1H, m), 4.22–4.13 (2H, m), 2.92–2.78 (2H, m), 1.23 (3H, t, *J* = 7.2 Hz); ¹³C NMR (100 MHz, CDCl₃) δ 170.1, 161.8, 146.8, 145.3, 130.3, 123.6, 72.1, 62.2, 54.6, 40.1, 14.0 p.p.m.; HRMS (*m/z*): [M + H]⁺ calcd. for: C₁₁H₁₁N₂O₅, 279.0975; found, 279.0978.

(2*S*,3*R*)-2-amino-3-hydroxy-4-(4-nitrophenyl)butanoic acid (**2**) was synthesized according to a literature procedure⁵⁶. Starting from 720 mg of **11** we obtained 570 mg of **2**. Yield: 92%. ¹H NMR (400 MHz, D₂O) δ 8.31 (2H, d, *J* = 8.5 Hz), 7.63 (2H, d, *J* = 8.5 Hz), 4.44 (1H, m), 3.84 (1H, m), 3.27 (1H, dd, *J*₁ = 14.0 Hz, *J*₂ = 3.2 Hz), 3.05 (1H, dd, *J*₁ = 14.0 Hz, *J*₂ = Yield: 92%. 1.6 Hz); ¹³C NMR (100 MHz, D₂O) δ 172.5, 146.6, 145.9, 130.5, 123.9, 70.3, 59.2, 39.8 p.p.m. HRMS (*m/z*): [M + H]⁺ calcd. for: C₁₀H₁₁N₂O₆, 241.0819; found: 241.0801. NMR data are in agreement with published values for the (2*S*,3*R*)-enantiomer⁴³.

(2*S*,3*R*)-2-amino-3-hydroxy-4-(4-aminophenyl)butanoic acid (**8**) was synthesized according to a literature procedure⁵⁴. Starting from 25 mg of **2** we obtained 7 mg of **7** as a yellow solid. ¹H NMR (400 MHz, D₂O) δ 7.37 (2H, d, *J* = 8.4 Hz), 7.28 (2H, d, *J* = 8.1 Hz), 4.39 (m, 1H), 4.05 (1H, d, *J* = 4.0 Hz), 3.01 (1H, dd, *J*₁ = 14.1 Hz, *J*₂ = 4.3 Hz), 2.82 (1H, dd, *J*₁ = 9.9 Hz, *J*₂ = 14.2 Hz); ¹³C NMR (100 MHz, D₂O) δ 170.35, 138.28, 130.86, 128.44, 123.17, 69.71, 57.16, 38.77 p.p.m. HRMS (*m/z*): [M + H]⁺ calcd. for: C₁₀H₁₃N₂O₅, 211.1077 [M + H]⁺; found, 211.1082.

The enzymatic synthesis of **2** was achieved by scaling up (to 2 ml) the analytical conditions for the ObAG discontinuous assay described above. The reaction was quenched after 2 h by the addition of MeOH (2 ml) and the solution was concentrated under reduced pressure. The resulting crude product was repeatedly subjected to SPE for further purification and to remove excess buffer. Elution with 25% MeOH yielded 2 (1.8 mg, 38%). NMR data were consistent with those of the synthetic reference standard. [α]_D²⁰ + 48°, (*c* = 0.18, H₂O) (Literature⁴³ [α]_D²⁰ + 50°, (*c* = 0.18, H₂O)).

Phylogenetic analyses. Amino acid sequences of *t*-TAs and SHMTs were obtained from a previous phylogenetic study⁵⁰ and were combined with additional *t*-TA and SHMT sequences for enzymes which have been characterized or described in the literature, and were obtained from the National Center for Biotechnology Information (NCBI) GenBank database⁵⁷ and the Protein Data Bank (PDB)⁵⁸. BLASTP (Basic Local Alignment Search Tool)⁵⁹ searches for *t*-TAs or SHMTs involved in natural product biosynthesis were performed to identify enzymes from these families associated with specialized metabolism. A selection of 10 amino acid sequences for 3-amino-5-hydroxybenzoic acid synthases (AHBAs) described in the literature were obtained to function as an outgroup for phylogenetic analysis as they have been shown to share a recent common ancestor of *t*-TAs and SHMTs⁶⁰ (Source organisms and amino acid sequence GenBank accession numbers used are reported in Supplementary Table 4). Sequences were initially aligned (all alignments performed with default settings) using ClustalX2 (ref. 61), before manual trimming of sequences at N- and C-termini to remove aberrant sequences (for example, histidine tags) that might interfere with the alignment. Trimmed sequences for Fig. 5 are reported in Supplementary Note 3. Several iterations of alignment and trimming were repeated with different degrees of trimming to ensure that the final tree was relatively consistent and robust (Supplementary Figs 15 and 16). Among the amino acid sequences used, FTase is unique in possessing a C-terminal phosphate-binding domain, and this was also trimmed (Supplementary Fig. 19). MUSCLE⁶², in addition to ClustalX2, was trialled for initial alignment before trimming (Supplementary Fig. 10). Trimmed sequences were finally re-aligned with T-Coffee⁶³, and phylogenetic tree inference was performed using maximum likelihood/rapid bootstrapping under the GTR model using RAxML-HPG BlackBox (8.2.8)⁶⁴ via the CIPRES Science Gateway portal⁶⁵. The JTT Protein Substitution Matrix was used and all other parameters were set to default values.

Data availability. The authors declare that the data supporting the findings reported in this study are available within the article and the Supplementary Information, or are available from the authors on reasonable request. New nucleotide sequence data have been deposited in NCBI GenBank under the accession codes as follows: 1 BGC (KX931446); pTS1 (KX931445); pME3087 (KX931444).

References

- Makris, T. M., Chakrabarti, M., Münck, E. & Lipscomb, J. D. A family of diiron monooxygenases catalysing amino acid beta-hydroxylation in antibiotic biosynthesis. *Proc. Natl Acad. Sci. USA* **107**, 15391–15396 (2010).
- Cryle, M. J., Meinhart, A. & Schlichting, I. Structural characterization of OxyD, a cytochrome P450 involved in beta-hydroxytyrosine formation in vancomycin biosynthesis. *J. Biol. Chem.* **285**, 24562–24574 (2010).
- Zweierink, M. M., Edison, A. M., Wells, G. B., Pinto, W. & Lester, R. L. Characterization of a novel, potent, and specific inhibitor of serine palmitoyltransferase. *J. Biol. Chem.* **267**, 25032–25038 (1992).
- Gulder, T. A. & Moore, B. S. Salinosporamide natural products: Potent 20S proteasome inhibitors as promising cancer chemotherapeutics. *Angew. Chem. Int. Ed. Engl.* **49**, 9346–9367 (2010).

- Goldstein, D. S. *t*-Dihydroxyphenylserine (*t*-DOPS): a norepinephrine prodrug. *Cardiovasc. Drug Rev.* **24**, 189–203 (2006).
- Kobayashi, J., Nakamura, M., Yamashita, & Kobayashi, S. Catalytic enatio- and diastereoselective aldol reactions of glycine-derived enolate with aldehydes: an efficient approach to the asymmetric synthesis of anti-beta-hydroxy-alpha-amino acid derivatives. *J. Am. Chem. Soc.* **126**, 9192–9193 (2004).
- Seiple, L. B., Mercer, J. A. M., Sussman, R. J., Zhang, Z. & Myers, A. G. Stereoccontrolled synthesis of syn-beta-hydroxy-alpha-amino acids by direct aldolization of pseudoephedrine glycinamide. *Angew. Chem. Int. Ed. Engl.* **53**, 4642–4647 (2014).
- Trost, B. M. & Miede, F. Development of ProPhenol ligands for the diastereo- and enantioselective synthesis of beta-hydroxy-alpha-amino esters. *J. Am. Chem. Soc.* **136**, 3016–3019 (2014).
- Liu, G.-S., Zhang, G.-Q., Yuan, Y.-A. & Xu, H. Iron(II)-catalyzed intramolecular aminohydroxylation of olefins with functionalized hydroxylamines. *J. Am. Chem. Soc.* **125**, 3343–3346 (2013).
- Baik, S. H. & Yoshioka, H. Enhanced synthesis of *t*-threo-3,4-dihydroxyphenylserine by high-density whole-cell biocatalyst of recombinant *t*-threonine aldolase from *Streptomyces avermitilis*. *Biotechnol. Lett.* **31**, 443–448 (2009).
- Franz, S. E. & Stewart, J. D. Threonine aldolases. *Adv. Appl. Microbiol.* **88**, 57–101 (2014).
- Fesko, K. Threonine aldolases: perspectives in engineering and screening the enzymes with enhanced substrate and stereo specificities. *Appl. Microbiol. Biotechnol.* **100**, 2579–2590 (2016).
- Wells, J. S., Trejo, W. H., Principe, P. A. & Sykes, R. B. Obafuorin, a novel beta-lactone produced by *Pseudomonas fluorescens*. Taxonomy, fermentation and biological properties. *J. Antibiot.* **37**, 802–803 (1984).
- Tymiak, A. A., Culver, A. A., Malley, M. F. & Gougoutas, J. Z. Structure of obafuorin. An antibacterial beta-lactone from *Pseudomonas fluorescens*. *J. Org. Chem.* **50**, 5491–5495 (1985).
- Herbert, R. B. & Knaggs, A. K. The biosynthesis of the *Pseudomonas fluorescens* antibiotic obafuorin from *p*-aminophenylalanine and glycine (glyoxylate). *Tetrahedron. Lett.* **31**, 7517–7520 (1990).
- Herbert, R. B. & Knaggs, A. K. Biosynthesis of the antibiotic obafuorin from *p*-aminophenylalanine and glycine (glyoxylate). *J. Chem. Soc. Perkin Trans. 1*, 109–113 (1992).
- Herbert, R. B. & Knaggs, A. K. Biosynthesis of the antibiotic obafuorin from *D*-[U-¹³C] glucose and *p*-aminophenylalanine in *Pseudomonas fluorescens*. *J. Chem. Soc. Perkin Trans. 1*, 103–107 (1992).
- Herbert, R. B. & Knaggs, A. K. The biosynthesis of the antibiotic obafuorin from *p*-aminophenylalanine in *Pseudomonas fluorescens*. *Tetrahedron Lett.* **29**, 6353–6356 (1998).
- Kluge, A. F. & Petter, R. C. Acylating drugs: redesigning natural covalent inhibitors. *Curr. Opin. Chem. Biol.* **14**, 421–427 (2010).
- De Pascale, G., Nazi, L., Harrison, P. H. M. & Wright, G. D. beta-Lactone natural products and derivatives inactivate homoserine transacylase, a target for antimicrobial agents. *J. Antibiot.* **64**, 483–487 (2011).
- Weber, T. *et al.* antimash3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–W243 (2015).
- Voisard, C. *et al.* in *Molecular Ecology of Rhizosphere Microorganisms*. (eds O’Gara, F., Dowling, D. N. & Boesten, B.) 67–89 (VCH, 1994).
- Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T. & Kado, C. I. Positive selection for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol.* **164**, 918–921 (1985).
- El-Sayed, A. K. *et al.* Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* **10**, 419–430 (2003).
- Walsh, C. T., Liu, J., Rusnak, F. & Sakaitani, M. Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway. *Chem. Rev.* **90**, 1105–1129 (1990).
- Drake, E. J., Nicolai, D. A. & Gulick, A. M. Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chem. Biol.* **13**, 409–419 (2006).
- He, J., Magarvey, N., Pirae, M. & Vining, L. C. The gene cluster for chloramphenicol biosynthesis in *Streptomyces venezuelae* ISP5230 includes novel shikimate pathway homologues and a monomodular non-ribosomal peptide synthetase gene. *Microbiology* **147**, 2817–2828 (2001).
- Blanc, V. *et al.* Identification and analysis of the genes from *Streptomyces pristinaespiralis* encoding enzymes involved in the biosynthesis of the 4-dimethylamino-*l*-phenylalanine precursor of pristinamycin I. *Mol. Microbiol.* **23**, 191–202 (1997).
- Mavrodi, D. V. *et al.* A seven-gene locus for synthesis of phenazine-1-carboxylic acid by *Pseudomonas fluorescens* 2-79. *J. Bacteriol.* **180**, 2441–2548 (1998).
- He, J. & Hertweck, C. Biosynthetic origin of the rare nitroaryl moiety of the polyketide antibiotic aureothin: involvement of an unprecedented N-oxygenase. *J. Am. Chem. Soc.* **126**, 3694–3695 (2004).

31. Choi, Y. S., Zhang, H., Brunzelle, J. S., Nair, S. K. & Zhao, H. *In vitro* reconstitution and crystal structure of *p*-aminobenzoate *N*-oxygenase (AurF) involved in aureothin biosynthesis. *Proc. Natl. Acad. Sci. USA* **105**, 6858–6863 (2008).
32. Knoot, C. J., Kovaleva, E. G. & Lipscomb, J. D. Crystal structure of CmlI, the arylamine oxygenase from the chloramphenicol biosynthetic pathway. *J. Biol. Inorg. Chem.* **21**, 589–603 (2016).
33. Hu, L., Chanco, E. & Zhao, H. CmlI is an *N*-oxygenase in the biosynthesis of chloramphenicol. *Tetrahedron* **68**, 7651–1654 (2012).
34. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362–W367 (2011).
35. Quadri, L. E. N., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T. Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-confirming siderophore mycobactin. *Chem. Biol.* **5**, 631–645 (1998).
36. Felnagle, E. A. *et al.* MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry* **49**, 8815–8817 (2010).
37. Baltz, R. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **38**, 1747–1760 (2011).
38. Heemstra, J. R., Walsh, C. T. & Sattely, E. S. Enzymatic tailoring of ornithine in the biosynthesis of the *Rhizobium* cyclic trihydroxamate siderophore vibactin. *J. Am. Chem. Soc.* **131**, 15317–15329 (2009).
39. Chen, H., Hubbard, B. K., O'Connor, S. E. & Walsh, C. T. Formation of β -hydroxy histidine in the conformation of nikkomycin antibiotics. *Chem. Biol.* **9**, 103–112 (2002).
40. Herbst, D. A., Boll, B., Zocher, G., Stehle, T. & Heide, L. Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylation enzymes. *J. Biol. Chem.* **288**, 1991–2003 (2013).
41. Gehring, A. M., Bradley, K. A. & Walsh, C. T. Enterobactin biosynthesis in *Escherichia coli*: Isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* **36**, 8495–8503 (1997).
42. Fu, C. *et al.* Solving the puzzle of one-carbon loss in ripostatin biosynthesis. *Angew. Chem. Int. Ed. Engl.* **56**, 2192–2197 (2017).
43. Pu, Y., Lowe, C., Sailer, M. & Vederas, J. Synthesis, stability and antimicrobial activity of (+)-obafuorin and related β -lactone antibiotics. *J. Org. Chem.* **59**, 3642–3655 (1994).
44. Lowther, J., Beattie, A. E., Langridge-Smith, P. R. R., Clarke, D. J. & Campopiano, D. J. *l*-Penicillamine is a mechanism-based inhibitor of serine palmitoyltransferase by forming a pyridoxal-5'-phosphate-thiazolidine adduct. *Med. Chem. Commun.* **3**, 1003–1008 (2012).
45. Chen, D. & Frey, P. A. Identification of lysine 346 as a functionally important residue for pyridoxal-5'-phosphate binding and catalysis in lysine 2,3-aminomutase from *Bacillus subtilis*. *Biochemistry* **40**, 596–602 (2001).
46. Kadi, N. & Challis, G. L. Siderophore biosynthesis: a substrate specificity assay for nonribosomal peptide synthetase-independent siderophore synthetases involving trapping of acyl-adenylate intermediates with hydroxylamine. *Methods Enzymol.* **458**, 431–457 (2009).
47. Meyer, S. *et al.* Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of NRPS. *Cell Chem. Biol.* **23**, 462–471 (2016).
48. Deng, H., Cross, S. M., McGlinchy, R. P., Hamilton, J. T. & O'Hagan, D. *In vitro* reconstituted biotransformation of 4-fluorothreonine from fluoride ion: application of the fluorinase. *Chem. Biol.* **15**, 1268–1276 (2008).
49. Barnard-Britton, S. *et al.* Amalgamation of nucleosides and amino acids in antibiotic biosynthesis: discovery of an *l*-threonineuridine-5'-aldehyde transaldolase. *J. Am. Chem. Soc.* **134**, 18514–18517 (2012).
50. Liu, G. *et al.* Evolution of threonine aldolases, a diverse family involved in the second pathway of glycine biosynthesis. *J. Mol. Evol.* **80**, 102–107 (2015).
51. Zelyas, N. J., Cai, H., Kwong, T. & Jensen, S. E. Alanylclavam biosynthetic genes are clustered together with one group of clavulanic acid biosynthetic genes in *Streptomyces clavuligerus*. *J. Bacteriol.* **190**, 7957–7965 (2008).
52. Mauchline, T. H. *et al.* Analysis of *Pseudomonas* genomic diversity in take-all infected wheat fields reveals the lasting impact of wheat cultivars on the soil microbiota. *Environ. Microbiol.* **17**, 4764–4778 (2015).
53. Kikuchi, G., Motokawa, Y., Yoshida, T. & Hiraga, K. Glycine cleavage system: reaction mechanisms, physiological significance, and hyperglycemia. *Proc. Jpn Acad. Ser. B Phys. Biol. Sci.* **84**, 246–263 (2008).
54. Herbert, R. B., Wilkinson, B. & Ellames, G. J. Preparation of (2*R*,3*S*)- β -hydroxy- α -amino acids by use of a novel *Streptomyces* aldolase as a resolving agent for racemic material. *Can. J. Chem.* **72**, 114–117 (1994).
55. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
56. Rao, M. N., Holkar, A. G. & Ayyangar, N. R. A novel transacylation method for the synthesis of α -*N*-acyl- β -lactones; application to (\pm)-diacetyllobafuorin and (+)-SQ 26,517. *J. Chem. Soc. Chem. Commun.* **1991**, 1007–1008 (1991).
57. Clark, K., Karsch-Mizrachi, L., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).
58. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
59. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
60. Contestabile, R. *et al.* *l*-threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase—a subgroup of strictly related enzymes specialized for different functions. *Eur. J. Biochem.* **268**, 6508–6525 (2001).
61. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
62. Edgar, R. C. MUSCLE: a multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
63. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
64. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
65. Miller, M. A., Pfeiffer, W. & Schwartz, T. In *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 1–8 (IEEE, New Orleans, LA, USA, 2010).

Acknowledgements

This work was supported by the BBSRC via Institute Strategic Programme Grant BB/J004561/1 to the John Innes Centre (JIC). The authors declare no conflicts of interest. We thank Dr Jacob Malone (JIC) for the supply of pME3087 and his invaluable assistance and expertise in *Pseudomonas* genetics. Plasmid pJH10 was a kind gift of Professor Chris Thomas and Dr Joanne Hothersall (University of Birmingham, UK). We also thank Dr Lionel Hill and Dr Gerhard Saalbach (JIC) for their excellent metabolomics support. Genome sequencing was undertaken at the Earlham Institute (Norwich, UK).

Author contributions

B.W. and T.A.S. designed the research. T.A.S., D.H. and B.W. wrote the manuscript. T.A.S. performed detailed bioinformatics and phylogeny analysis, carried out the mutational analysis, purified proteins and ran biochemical experiments. D.H. carried out the ObaG NMR experiments and isolated enzymatically synthesized compounds. D.H. and B.W. synthesized chemical standards. Z.Q. isolated obafuorin and analogues and carried out NMR analysis.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Scott, T. A. *et al.* An *l*-threonine transaldolase is required for *l*-threo- β -hydroxy- α -amino acid assembly during obafuorin biosynthesis. *Nat. Commun.* **8**, 15935 doi: 10.1038/ncomms15935 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

Figure 1.2. NPs from *Streptomyces coelicolor* A3(2). The colour-based phenotypes associated with the production of some *S. coelicolor* NPs made it an ideal model organism for biosynthetic studies. (a) Clockwise from top: WT *S. coelicolor*, *S. coelicolor* Δ act Δ red (actinorhodin and prodigiosin BGCs knocked out), and *S. coelicolor* Δ act. The WT strain is producing actinorhodin (blue) and the Δ act strain can be seen to be producing prodigiosin (pink/red). (b) *S. coelicolor* Δ act Δ cda Δ red (Δ cda = calcium-dependent antibiotic BGC knocked out). The yellow metabolite diffusing from the streaked strain is coelimycin. (c) Diversity of NPs encoded in the *S. coelicolor* A3(2) genome (coelibactin is a predicted structure). Many were only characterised after their BGCs had been identified as they were not produced/detectable using standard culture conditions. Plate images are courtesy of Dr. Juan-Pablo Gomez-Escribano in the Wilkinson group.

Figure 1.5. Illustration of NRPS assembly line biosynthetic logic. (a) Chain elongation on a hypothetical NRPS assembly line. (b) The condensation catalysed during NRPS biosynthesis. Nucleophilic attack of the downstream aminoacyl-S-PCP amine on the upstream peptidyl-S-PCP leads to the formation of an amide bond. (c) Reaction catalysed by NRPS epimerisation (E) domains. The E domain epimerises the upstream L-peptidyl-S-PCP donor to a D-peptidyl-S-PCP prior to condensation. These two steps can be catalysed by the sequential action of independent E and C domains, or both reactions can be catalysed by C/E didomains. (d) Following condensation, cyclisation (Cy) domains catalyse the attack of nucleophilic cysteine, serine or threonine side chains on the 'upstream' carbonyl before catalysing subsequent dehydration of the resulting adduct to yield thiazole or oxazole rings.

Figure 3.5. BGC architecture and proposed biosynthetic pathway for the biosynthesis of obafluorin (1). This pathway is based on the hypothetical pathway for **1** proposed by Herbert and Knaggs (1988; 1990; 1992a; 1992b) and in silico analysis described in this Chapter. PCSs *obaA-N* (~19 kb) are labelled and coloured according to anticipated functions during the biosynthesis of **1**, with putative operons indicated with black arrows beneath the gene cluster. Orfs12345678 represent PCSs that could contribute to **1** production, but which are not considered essential. The domain architectures of the ArCP *ObaK* and NRPS *Obal* are represented, and the anticipated mechanism of TE-catalysed peptide release and β -lactone ring closure is also illustrated (described further in Chapter 7).

Figure 4.4. BGC architecture and revised pathway for the biosynthesis of obafluorin (1) based on mutagenetic data. This pathway is based on mutagenetic data described in this Chapter. PCSs obaA-N are labelled and coloured according to related biosynthetic functions, with putative operons indicated with black arrows beneath the gene cluster. The domain architectures of the ArCP ObaK and NRPS ObaI are represented, and the anticipated mechanism of TE-catalysed peptide release and β -lactone ring closure is also illustrated (described further in Chapter 7). **4** and **5** are derived from some element of the biosynthesis of **2**, and were ultimately identified as breakdown products of (4-NPA; **11**). **4** and **5** accumulate when either the biosynthesis of **6**, or the assembly of **1**, was disrupted. Compounds **3** and **9** were used in chemical feeding experiments in this work, but were shown not to be intermediates in the biosynthesis of **1**.

Figure 5.8. Experimental confirmation that ObaG binds PLP. (a) UV/Vis spectrum of 20 μM His₆-ObaG (blue) showing peaks at 340 nm and 390 nm, representing the enolimine and ketoenamine forms of the PLP cofactor respectively. Addition of 10 mM L-penicillamine (red) led to formation of a thiazolidine adduct (340 nm peak) and concomitant loss of the 390 nm peak following 30 min incubation. (b) UV/Vis spectrum of 20 μM His₆-ObaG (blue). Addition of 1 mM NaBH₄ (red) led to formation of a covalent His₆-ObaG-PLP adduct (330 nm peak) and concomitant loss of the 390 nm ketoenamine peak following 15 min on ice. (c) The enzyme is pink in solution (right), but following boiling (left), the pink colour is lost and instead a faint green/yellow colour can be seen, indicating release of the PLP cofactor. (d) L-penicillamine reacts with PLP via an external aldimine intermediate to form a thiazolidine adduct. (e) Reduction of the internal aldimine formed between the enzyme catalytic lysine residue and PLP to form a secondary amine adduct.

Figure 6.2. Maximum likelihood tree of L-TAs, SHMTs and L-TTAs. A set of 3-amino-5-hydroxybenzoate synthase (AHBAS) amino acid sequences serve as the outgroup. RAxML likelihood values at the root nodes for SHMT and L-TA cluster clades are annotated. The clade comprising the L-TTA ObaG characterised in this work has been expanded and annotated with RAxML likelihood values and the identity of amino acid sequences represented. Initial trimming was performed following alignment with ClustalX2, before final alignment with T-Coffee. Circles indicate PacT homologues (pink), VrtJ (yellow), FmoH and AmiS (blue) and FmoM (green).

Figure 6.3. Proposed biosynthetic pathways to alanylclavam, the clavamycins and valclavam. Purple – biosynthesis of alanylclavam in *S. clavuligerus* NRRL 3585; turquoise – biosynthesis of clavamycin A in *S. hygrosopicus* NRRL 15879; and dark blue – biosynthesis of valclavam in *S. antibioticus* Tü1718. The proposed reactions catalysed by OrfA-like GHMT homologues are highlighted in orange. We hypothesise that these enzymes are in fact L-TTAs that use L-threonine rather than glycine (as depicted) as a co-substrate with aldehydes **X** (alanylclavam and clavamycin biosynthesis) or **Y** (valclavam biosynthesis). The products of these reactions are 8-hydroxyalanylclavam and a 9-hydroxyalanylclavam intermediate, **Z**, respectively. All pathways share the same conserved steps to clavaminic acid. Jensen (2012) - Adapted with permission from Springer.

Figure 8.1. Illustration of the diverse bacterial targets for microbial NPs. A comprehensive review of the antibacterial NPs identified 54 distinct molecular targets of known antibacterials, including:

Membrane-associated targets - 1) lipid II; 2) undecaprenyl phosphate; 3) phosphatidylethanolamine; 4) Na-dependent NADH-quinone reductase; 5) translocase 1; 6) SecA–YEG complex; 7) type 1 signal peptidase; 8) type 2 signal peptidase; 9) WalK-WalR two-component system; 10) NADH oxidase; 11) menaquinone; 12) cytochrome bd; 13) membrane stability; 14) LPS

Cell wall-associated enzymes - 15) peptidoglycan transglycosylase; 16) penicillin-binding proteins; 17) D-alanine-D-alanine dipeptide; 18) UDP-N-acetylglucosamine-3-enolpyruvyl transferase

Targets associated with amino acid biosynthesis and metabolism - 19) L-histidinol phosphate aminotransferase; 20) N-acetylornithine transaminase; 21, isoleucyl-tRNA synthetase; 22, leucyl-tRNA synthetase; 23, tryptophanyl-tRNA synthetase; 24) prolyl-tRNA synthetase; 25) tyrosyl-tRNA synthetase; 26) threonyl-tRNA synthetase; 27) phenylalanyl-tRNA synthetase; 28) aspartyl-tRNA synthetase; 29) seryl-tRNA synthetase; 30) aspartate semialdehyde dehydrogenase; 31) alanine racemase; 32) glutamine synthase; 33) pyruvate dehydrogenase; 34) threonine synthase; 35) biotin aminotransferase; 36) homoserine-O-succinyltransferase; 37) pyruvate carboxylase; 38) peptide deformylase

Fatty acid biosynthesis - 39) FabF; 40) FabB; 41) acyl-CoA carboxylase; 42) FabD; 43) FabG; 44) FabI, individual metabolic enzymes (45) glucosamine-6-phosphate synthase; 46) deoxyxylulose phosphate reductoisomerase; 47) ADC synthase

Components of macromolecular machinery - 48) ClpP protease; 49) elongation factors; 50) ribosome; 51) transcription termination factor Rho; 52) RNA polymerase; 53) DNA gyrase; 54) DNA polymerase

Johnston et al. (2016) – Adapted with permission from the Nature Publishing Group.

Figure 8.10. Illustration of the mutasynthetic strategy to generate novel congeners of obafluorin (1) in the Δ obaL strain. 6 should no longer be produced in the Δ obaL strain, allowing the exogenous introduction of alternative benzyl carboxylic acids (coloured in burgundy). If the Obal A₁ domain is sufficiently promiscuous, it may accept these alternative substrates and assemble them together with 2 to create novel 1-like metabolites. Downstream-acting domains must also be able to tolerate these alternative substrates.